RICE UNIVERSITY

Reverse-Engineering Self-Organized Behavior in *Myxococcus xanthus* Biofilms

by

Haiyang Zhang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Oleg A. Igoshin, Chair
Assistant Professor, Department of Bioengineering, Rice University

Robert M. Raphael
Associate Professor, Department of Bioengineering, Rice University

Marek Kimmel
Professor of Statistics, Department of Statistics, Rice University

HOUSTON, TEXAS
December 2011

Abstract

Reverse-Engineering Self-Organized Behavior in *Myxococcus xanthus* Biofilms

by

Haiyang Zhang

Myxococcus xanthus (*M. xanthus*) is a gram-negative, rod-shaped soil-dwelling predatory bacterium. It can move on solid surfaces forming cooperative single-species biofilm in which various self-organizing patterns are observed. Under distinct environmental conditions, these bacteria can swarm outward, form travelling waves or aggregate into fruiting bodies as a result of diverse intercellular interactions, signaling and coordinated cell motility. *M. xanthus* colony actively expands when food is plentiful, but stops this under nutritional stress and thereafter aggregates into fruiting bodies where individual cells transform into spores. When in direct contact with their prey, *M. xanthus* cells form traveling cell–density waves called ripples to facilitate their predation. These patterns play an important role in maximizing *M. xanthus* adaption to the changing environment. While these phenomena have been studied using traditional experimental microbiology and genetics, recently it is becoming clear that system biology approach greatly complements traditional laboratory work. This thesis shows my effort to deepen the understanding of self-organization in microorganisms using statistical image processing techniques and agent-based modeling. Statistical image processing

results illustrate that aggregation into fruiting bodies is a highly non-monotonic yet spontaneous process without long-range signal transduction. The agent-based model of aggregation accurately reproduces the final steady states of an aggregation process but fails to reproduce the experimental dynamics. The agent-based modeling for predatory ripples quantitatively reproduces all observed patterns based on three simple experimentally observed rules: regular cellular reversals, side-to-side contact induced early reversals and refractory period after each cellular reversal. Moreover, the agent-based model predicts that predatory ripples speed up the swarm expansion into the prey region and keep individual *M. xanthus* cells in the prey region longer. These predictions are all quantitatively verified by experimental observations. The combination of statistical image analysis and agent-based modeling brings greater understanding of self-organizing patterns in *M. xanthus* and will be essential for further research on similar patterns in other microorganisms and higher organisms.

Acknowledgments

I would like to express my sincere gratitude to my advisor, Dr. Oleg A. Igoshin. I have been really fortunate to have an advisor who gave me the freedom to explore, and the guidance to recover when my steps faltered. His creative idea, inspiring talks, understanding and patience is indispensible for my graduate study. I would also like to thank all other dissertation committee members: Dr. Robert M. Raphael, Dr. Marek Kimmel and the former member Dr. Marina Vannucci, for their time and expertise to better my work.

I would like to thank Dr. Heidi Kaplan's group from UT-Houston Medical School and Dr. Roy Welch's group from Syracuse University. As collaborators, they not only provided valuable experimental data for my research but gave me heuristic ideas through many helpful discussions as well.

I would like to thank all the current and former members in Dr. Igoshin's group: Christian Ray, Jatin Narula, Abhinav Tiwari, Srabanti Chaudhury, Chunyan Xie, Baris Hancioglu, and Pen Shi. I have had a wonderful graduate school experience with all of them and this dissertation would not be possible without help from them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Introduction to *M. xanthus* bacteria

M. xanthus is a gram-negative bacterium commonly found in soil [1-3]. The bacterium has a rod shaped cell body(Figure 1-1 A) approximately 5-7 μm long and 0.5-1 μm in diameter [4]. *M. xanthus* lives a complex and collaborative life-cycle (Figure 1-2) and forms various self-organized patterns such as traveling cell-density waves (ripples) during predation and starvation (Figure 1-1 C)[5-11], fruiting-body aggregates (Figure 1-1 B) during starvation [3, 12-14], and protrusions and rafts under nutrient-rich conditions (Figure 1-1 D)., *M. xanthus* predates on other microorganisms in a highly cooperative manner by forming a single-species biofilm and feeds by metabolizing insoluble macromolecules [9, 11, 15]. The various self-organizing behaviors make *M. xanthus* a perfect model system in scientific research of cellular differentiation, social interactions and evolution [3, 16-18].

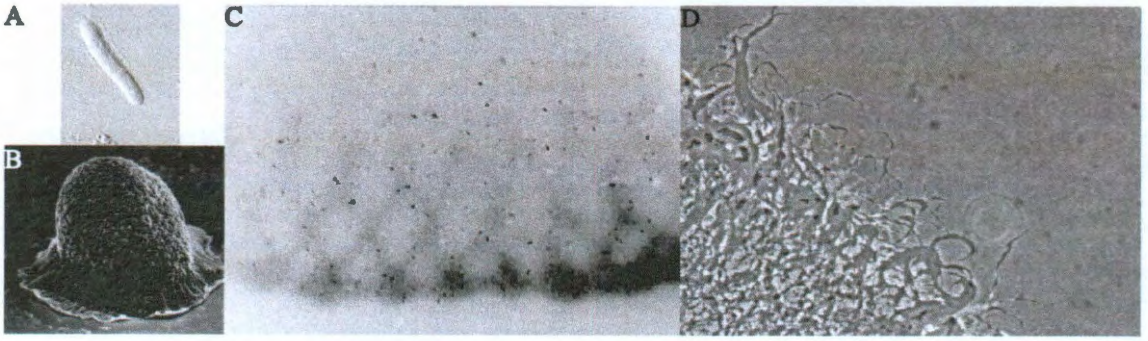


Figure 1-1 Experimental images of *M. xanthus* cells and swarms. (A) shows a single *M. xanthus* cell under a microscope (image adapted from [19]). (B) shows a mature fruiting body that contains up to a million cells (adapted from [14]). (C) shows the ripple pattern (image adapted from [8]). (D) shows the *M. xanthus* colony expansion (image adapted from [20]).

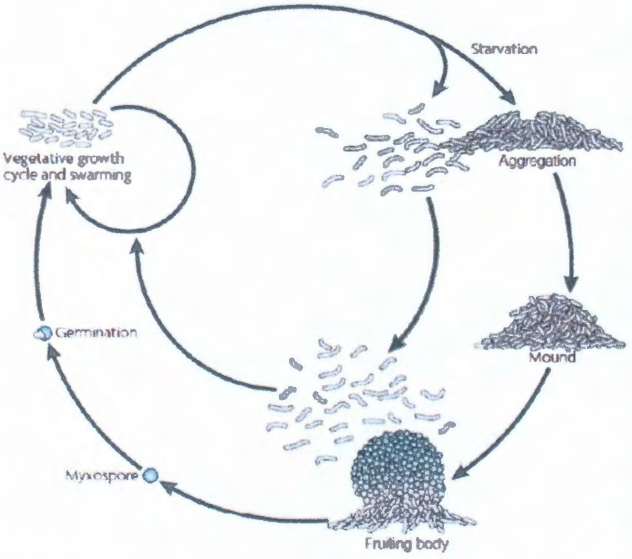


Figure 1-2 *M. xanthus* life-cycle (figure adapted from [5]). With favorable nutrient level, *M. xanthus* swarms grow and expand to search for additional resource. During starvation, swarms stop growing and *M. xanthus* cells start to aggregate to form fruiting bodies. Cells inside fruiting bodies form myxospores and peripheral cells maintain vegetative state. When the nutrient level recovers, the myxospores germinate. The fruiting bodies collapse and form new vegetative growing swarms.

Much effort has been devoted into the attempts to gain an understanding of the self-organizing patterns of *M. xanthus* swarms [7, 9, 21-41] by both experimental and theoretical approaches. Accurate characterization of these self-organizing patterns requires knowledge of mechanisms of cell motility, intercellular

interactions and signaling. Traditional experimental research accumulates valuable data and proposes some hypothetical mechanisms, but lacks means for quantitative and comprehensive explanations of the dynamics and physiological roles of the self-organization behaviors[3, 5, 6, 8, 24, 42-59]. System biology research, on the other hand, aims to provide an integrated view of these biological phenomena. Among various system biology methods, the agent-based modeling and statistical image processing emerge as approaches to study the collective motion patterns of microorganisms [23, 34, 36, 39, 41, 60-62]. In this thesis, I present my work on exploiting agent-based modeling and statistical image processing techniques to uncover the mysteries of self-organizing patterns observed in *M. xanthus* swarms.

1.1.1. Motility of *M. xanthus*

Unlike many other motile bacteria, *M. xanthus* cells lack flagella and cannot swim in liquids but nevertheless can move on solid surfaces at around 3-6 μm/min. This motility is often called gliding as it involves smooth translocation of a cell along its long axis [63]. Cells utilize two distinct and almost independent systems in its gliding motility [5, 44, 45, 64, 65]. One is the A-motility (Adventurous motility) system and the other is the S-motility (Social motility) system. As the name suggests, the A-motility system endows *M. xanthus* cells with the ability to move as individual cells whereas the S-motility system is involved in the motion of cell groups. Hodgkin and Kaiser isolated mutant strains that utilize only A-motility or S-motility to study the individual motility systems, as well as the synergy between them [44, 64, 65]. A further study of the influence of individual cell motility on swarming by Kaiser and

Crosby revealed that two motility systems are independent. The authors demonstrated that A-motile cells move at the same speed as wild type cells with no adjacent cells and S-motile cells move within a pack of cells at the same speed as wild type cells[66]. Furthermore, Kaiser and Crosby discovered that a wild-type *M. xanthus* colony has the largest expansion rate, S-motile colony has the smallest swarm expansion rate while A-motile colony has an expansion rate in-between. Moreover, the rate of expansion of wild-type cells exceeds the sum of the rates of expansion of strains with only one motility form. This suggests that synergy must exist between the two motility systems resulting in swarming rate increase. Wild type and mutant strains have distinct swarming patterns [66]. The colony edge of the S-motile strain rarely contains isolated cells, the colony edge of the A-motile strain contains quite a lot of isolated cells, while the colony edge of the wild type strain is in between (Figure 1-3).

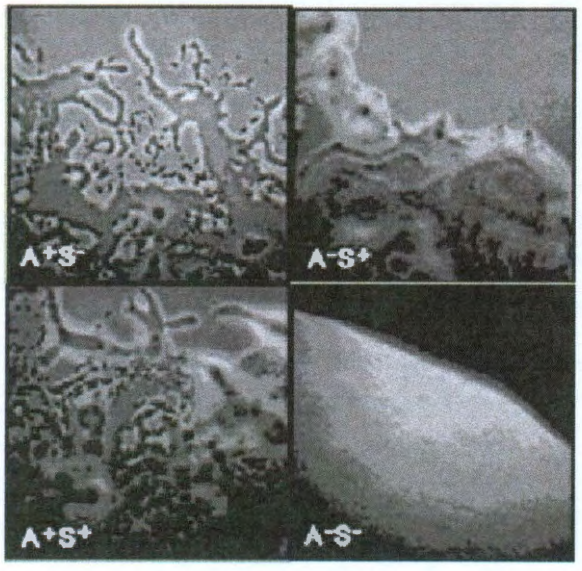


Figure 1-3 Distinct swarming patterns of different motility-mutated *M. xanthus* strains (adapted from [67]). A+S+ denotes wild type. A+S- denotes A-motile only. A-S+ denotes S-motile only. A-S- denotes defective in both A and S motility systems.

M. xanthus cells never constantly glide in one direction. Gliding direction changes not only due to random fluctuations of cell axis, but also due to periodical reversal of the polarity of the two cellular poles (leading pole becomes lagging and vice versa) every 6-10 minutes. In *M. xanthus*, control of the periodical cellular reversal is done by the *Frz* system containing seven *frz* genes [5, 68]. The *Frz* system was first discovered in mutants defective in development under starvation conditions[46]. Most of the *frz* deletion mutant strains show rare cellular reversals yet *frzD* mutant strains show hyper cellular reversal phenotype with reversals every 2 min [5, 68]. Evidently the *Frz* system is required for formation of self-organizing patterns [5, 7, 8, 23, 46] and the reversal time is optimized for a maximum expansion rate of *M. xanthus* colony [40].

1.1.1.1. Mechanism of Adventurous motility

A-motility of *M. xanthus* cells was discovered almost thirty years ago by Hodgkin and Kaiser [44, 64, 65]. Up to date about 40 genes have been identified to be related to A-motility. However, the mechanism of A-motility is still not yet thoroughly understood. There are several A-motility models. Wolgemuth et al proposed that A-motile cells can secret slime through the pores at the rear end of the cell body Figure 1-4 A. The hydration of slime when mixed with water generates the force which propels the cell forward [55, 69]. The focal adhesion model was proposed by Mignot et al (Figure 1-4 B)[57, 58]. Recently, Nan et al introduced a helical rotor model to explain the mysterious A-motility of *M. xanthus* (Figure 1-4 C)[43]. The details of these models are described below.

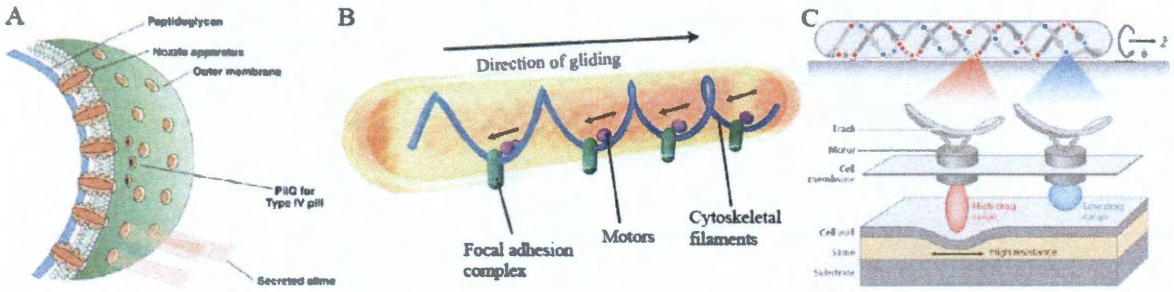


Figure 1-4 Illustration of different A-motility model. (A) Slime extrusion model (adapted from [55]). *M. xanthus* cells secrete slime through the nozzle at lagging poles producing force associated with slime hidration. (B) Focal adhesion model (adapted from [70]). (C) Helical rotor model (adapted from [70]).

As cells glide on a surface, they secrete slime at the lagging pole of their cell bodies, leaving a trail for other cells to follow (Figure 1-4 A). The slime contains polysaccharides with other secreted material. Studies of Wolgemuth et al provided supporting evidence that the slime extrusion engine at the rear end of a *M. xanthus* cell body can indeed provide enough force to maintain the speed observed in experiments [55]. However, contradictory experimental observations reported that the A-motility engine is distributed along the cell body and the distribution might not be necessarily uniform [71, 72]. Slime extrusion might not be the mechanism for A-motility after all.

An alternative A-motility model has been proposed by Mignot et al [57, 58]. The authors proposed the focal adhesion model based on the observation that *AglZ* (an A-motility protein in the cytoplasm) clumps and remains stationary relative to the substrate surface as A-motile cells glide over. According to Mignot et al, some motor proteins moving along helical cytoskeletal filaments can attach to focal

adhesion complexes sticking to the extracellular matrix and provide the propulsion force for *M. xanthus* cells locomotion [57, 58] (Figure 1-4 B). This model is relatively simple since it is based on one motility protein, *AglZ*. It is notable that the focal adhesion complexes can break the integrity of cells since they penetrate the cell body to stick to the extracellular matrix and constantly cut the cell envelope as the cell moves on[70].

Following Mignot et al's study, Nan et al proposed a helical rotor model for *M. xanthus* A-motility. In this model, homologs to MotAB, the protein-motive-force-drive proteins traveling along a looped cytoskeletal filament, are the power source. A-motility proteins *AgmU* and *AglZ* belong to larger protein complexes that push against the substrate surface as they are carried by the motor proteins. When the motors move through the ventral side of the cell body, those motors that carry high drag force cargoes slow down and form the clusters (Figure 1-4 C). Then the motion of aggregated motors with cargos can generate enough force to power the gliding motility[43, 70].

Overall, the A-motility system has great evolutionary significance since it allows cells to move individually and leave the swarm to probe the environment and quickly occupy a food source.

1.1.1.2. Mechanism of Social motility

M. xanthus cells with only S-motility (lacking A-motility genes) show extremely low motion of isolated cells and regular motility for a group of packed

cells with the same speed as wild type cells [24]. S-motility is powered by the retraction force of Type IV pili located at the front end of *M. xanthus* cell bodies (Figure 1-5) [73, 74]. Pili extend from the leading cell pole and bind to polysaccharides, either on another *M. xanthus* cell or the substrate surface [53]. Then the pili retract and pull the cell forward. When a cell reverses, the pili extend from the new leading pole (the previous lagging pole) to maintain the cellular movement in the new direction.

Figure 1-5 Microscopic image of pili extending from the leading pole of *M. xanthus* cells (adapted from [75]). White scale bar is 2μm.

There are still questions about how the *M. xanthus* cells on the leading edge of an expanding colony move forward, since there are no more cells in front of them. Researchers have observed that a colony of S-motile strain expands in a speed much larger than non-motile cells [24]. Though the pili model provides the possibility that S-motile cells can use their pili to attach to substrate surface, this hypothesis seems contradictory to the observation that single S-motile cells without neighboring cells do not move at all.

As a cooperative species, *M. xanthus* requires a social life for survival as its predation and development need collective motion of cells. S-motility keeps a tight

connection among *M. xanthus* cells and is responsible for the collective motion during development.

1.1.1.3. *Frz* system regulated periodical cell reversals

There is a multigene system that controls the regular cellular reversals of *M. xanthus* cells. The system includes seven genes named *frzA, frzB, frzC, frzD, frzE,,frzF,* and *frzG.* Mutation of any of these genes results in irregular cellular reversals of *M. xanthus* cells. Except for the *frzD* mutant, *M. xanthus* cells with mutation of other genes show extremely low reversal frequency (reversal period~2hour)[76]. Mutation in *frzD,* on the other hand makes cells reverse more frequently than normal (~2.2minutes mutant vs. 6-8 minutes wild type) [76]. The products of the *frz* genes form a signal transduction system in cytoplasm. Like chemotaxis pathways in *E. coli*, the system is a modified two-component system including *FrzCD* (products of *frzC* and *frzD* genes), the chemoreceptor; *FrzE*, the histidine kinase; *FrzA* and *FrzB*, the coupling proteins; and a dual response regulator protein *FrzZ*. The system also contains methyltransferase *FrzF* and methyesterase *FrzG*. *FrzCD* forms helix filaments spanning the entire cell length. Cellular reversals of *M. xanthus* correlate with the methylation of *FrzCD* [77]. Hyper-reversing cells have high demethylated *FrzCD* and hypo-reversing cells have high methylated *FrzCD* [77].

Controlled by the *Frz* chemosensory system, *M. xanthus* cells periodically reverse their moving directions by switching the polarity of the leading end and rear end. The polarities of both A and S motility systems change. During cell reversal the type IV pili of the S-motility system are disassembled at the previous leading pole

and reassembled at the current leading pole of an *M. xanthus* cell. A similar process is observed for the A-motility system. A-motility proteins have been observed assembling in clusters at the new leading cell pole when an *M. xanthus* cell reverses [42]. Controlling the reversal periods is the way *M. xanthus* cells do a bias random walk so that they can move away from harmful conditions. When an attractant presents, *M. xanthus* increases its reversal period and when a repellent presents, *M. xanthus* decreases its reversal period[5]. This is highly similar to *E. coli*'s chemotaxis behavior in that when an attractant presents *E. coli* does "run" often and when a repellent presents *E. coli* does "tumble" more often[5].

1.1.2. Intercellular signaling

As highly cooperative species, *M. xanthus* cells exchange various signals with their neighbors during their collective motility. C-signal is a short-range signal that is exchanged through end-to-end contacts on the cell surfaces [6-8, 31, 78] (Figure 1-6 A). The transduction of the C-signal increases the reversal frequency of *M. xanthus* cells and is required for formation of ripples during their developmental process [6, 7, 36, 56]. Another contact signal is activated during the predation of *M. xanthus* on other microorganisms[59]. A recent study tracking localization of the chemoreceptor *FrzCD* revealed that when two long axis parallel *M. xanthus* cells are close enough to each other side-to-side, their clusters *FrzCD* receptors aligned (Figure 1-6 B) [59]resulting in the reversal of one of the cells.

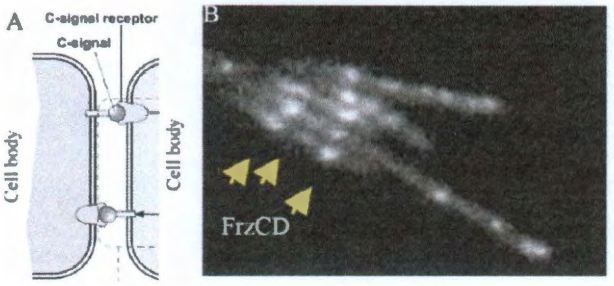


Figure 1-6 (A) Schematic representation of head-to-head C-signaling (adapted from [56]). (B) Experimental flourescent image that shows the alignment of *FrzCD* of *M. xanthus* cells with side-to-side contact (adapted from [59]).

1.1.3. *M. xanthus* self-organizing patterns

The ability of cells to self-organize into complex dynamical structures during development remains one of the most intriguing fundamental principles in modern biology [29, 79-81]. Even as life sciences stand firmly in the post-genomic era and the genomes of many organisms have been sequenced, researchers lack a complete mechanistic understanding of the interactions that ultimately determine the size and shape of an organism, define its segments and polarity, and regulate the production and release of intercellular signals. Albeit less complex, the formation, maturation and release of bacterial biofilms is also a form of multicellular development [82-85]. As such, the dynamic organization of the biofilms formed by *M. xanthus* is dependent on the ability of the cells to sense, integrate, and respond to a variety of intercellular and environmental cues that coordinate surface motility [86-89].

M. xanthus has a complicated yet highly cooperative life cycle [50, 90]. When nutrients are rich, *M. xanthus* swarm grows and spreads vegetatively [12, 38, 91]. When a prey is present, *M. xanthus* swarms can form traveling high cell density waves termed predatory ripples [9, 10]. When nutrient level is low, the *M. xanthus* swarms first form the traveling high cell density waves termed developmental

ripples [8, 35, 92]. After a certain time under starvation, *M. xanthus* cells start their developmental program that leads the cells to aggregate into fruiting bodies[21, 28], where cells develop into spores [13, 93]. Once the environment recovers, e.g. the nutrient level returns to normal, the spores germinate and form new vegetative swarms again [26, 94], which completes the life cycle.

1.1.3.1. Ripple patterns

Ripples of *M. xanthus* are traveling high cell density waves occurring during both development and predation[7, 8, 10, 11]. Starting from a uniform distribution of cells, *M. xanthus* cells gradually form dense traveling bands over a period of several hours. Wave crests usually have 3-8 times higher cell densities than troughs and can be observed traveling at the same speed as the individual cells inside them. The high density wave crests travel back and forth and reflect off each other when counter-moving wave crests collide. During development, contact mediated C-signal is transduced between *M. xanthus* cells making end-to-end contact, which results in the ripple patterns termed developmental ripple [7, 8]. Ripple patterns termed predatory ripples form during predation. There are two possible signaling mechanisms behind predatory ripples. One is that during predation, the extracellular digestive enzymes secreted by *M. xanthus* can cause its own cell lysis, which triggers C-signaling transduction and directs some of *M. xanthus* cells into the developmental program. My work described in this thesis provides an alternative side-to-side signaling other than C-signal can also direct cells into ripple patterns.

Ripples under the two conditions have great evolutionary significance. The developmental ripples may help pinpoint the position of fruiting bodies and the predatory ripples may facilitate the predation efficiency.

1.1.3.2. Aggregation into fruiting bodies

A swarm placed on nonnutritive starvation agar exhibits a different set of behaviors, collectively referred to as development[3]. When development starts, swarm expansion stops and then the entire swarm appears to contract as movement is redirected inward. Over a period of several hours, thousands of cells accumulate at certain locations within the swarm, and once there, they arrange themselves into multicellular structures called aggregates. Aggregation is a highly non-monotonic process. The number of aggregates peaks in about 5 hours and then gradually decreases. During the same time, individual aggregates can experience the fates of growth, dispersal, merger and splitting [22, 23]. It then takes less than a day before the aggregates fully mature to fruiting bodies. During the maturation of fruiting bodies, cells inside the aggregates start to transform into spores that are resistant to the harsh environment. Some cells outside the aggregates either die to form a protective cover while others remain in a vegetative state and monitor the surrounding environment.. At this point, an aggregate is considered to have matured into a fruiting body. Each fruiting body is dome--shaped and 0.1 mm in diameter [14, 37] containing up to millions of cells (spores) [21, 28]. When environmental conditions recover, e.g. the nutrient level returns to normal, the spores inside

fruiting bodies geminate. The fruiting bodies collapse and form new swarming colonies.

1.1.3.3. Swarm expansion

Under favorable conditions such as rich nutrients or the presence of prey, an *M. xanthus* colony expands as individual cells move actively in coordination with each other. *M. xanthus* cells move out from the swarm edge in multicellular projections called flares. On a smooth nutrient agar surface, the flares radiate in all directions. As they expand and merge with each other, they form a contiguous population that increases the area of the swarm. In this way, a circular swarm expands radially with a relatively even and symmetrical distribution of cells. The expansions of *M. xanthus* colony of various motility mutants have been studied by Kaiser et al [24]. Except for motility double mutant (lacking both A and S motility), the expansion rate positively correlates with the initial colony cell density until it saturates: the initial colony cell density reaches a threshold[24]. Different motility mutant strains have distinct patterns in their swarm expansion, which has made swarm expansion a useful phenotype for characterization of the different motility mutations of *M. xanthus*.

1.2. Modeling *M. xanthus* self-organization

M. xanthus self-organization patterns cannot be explained solely by available experimental approaches; thus, mathematical modeling has been used extensively to gain a deeper understanding of these emergent behaviors[7, 32, 34, 36, 39-41,

95]. These mathematical models have shown that a system biology approach is an excellent complement to classic biology research in uncovering the mysteries of microorganisms' social interactions. In 2001 Igoshin et al developed a mean field model that revealed the mechanism of developmental rippling in *M. xanthus* [7]. The model emphasized the interactions between cells observed in experiments and provided a quantitative characterization of developmental ripple pattern, indicating that the wavelength is twice the distance the *M. xanthus* cells travel before they reverse. The results demonstrated that the developmental ripple in *M. xanthus* is different from wave patterns observed in other social organisms such as *Dictyostelium discoideum*. As a next step, Igoshin et al developed a model to describe another self-organizing pattern in *M. xanthus*: aggregation into fruiting bodies. The model was based on traffic jamming of cells in high cell density regions [32]. Later, a lattice-gas cellular automata model was developed by Sozinova et al. describing the aggregation into fruiting bodies [34]. The model reproduced some patterns of the aggregation process and showed that traffic jams could be responsible for aggregation initialization[34]. Agent-based models were then introduced into modeling the self-organizations of *M. xanthus* [36, 39, 40, 62, 95], including. Anderson et al's 1-D individual based model following the work of Igoshin et al in 2001[62], Sliusarenko et al's agent-based model for developmental rippling and aggregation into fruiting bodies [36, 39], and Wu et al's agent-based model to study the swarming of *M. xanthus* colonies[40, 95].

1.2.1. Cell-density based models

Sager and Kaiser proposed that head-to-head C-signals between *M. xanthus* cells could lead to early cellular reversal[6]. Furthermore, they discovered that during rippling, cells are still travelling back and forth with wave crests[8] and the cell density waves are not unidirectional. The cells move back and forth and reflect off each other as opposing wave crests collide. In 2001, Igoshin et al. proposed that C-signals can shorten the cellular reversal period and is the key to symmetry breaking [7]. The Igoshin et al's mean field model assumes that each *M. xanthus* cell has a regular reversal period of 6-8 minutes. During starvation conditions, *M. xanthus* cells start to exchange C-signal with opposing neighbors near their leading poles. The exchange of C-signals leads to shortening the reversal period by a certain amount, and several C-signaling events can trigger the cellular reversal. *M. xanthus* cells have a refractory period after each cellular reversal during which they do not respond to C-signals. The mean field model quantitatively reproduced the rippling patterns and revealed how contact-mediated intercellular signal and cell motility together determine the social organization in *M. xanthus*. The results also demonstrated that the mechanism behind the rippling pattern did not relate to diffusive chemicals but rather involved cellular reversals regulated by the *Frz* system [7].

1.2.2. Cellular automata models

In the cellular automata models, space is defined as a lattice grid and individuals are defined by arrays of nodes on the lattice. In 2005, Sozinova et al

introduced a cellular automata model for the aggregation process of *M. xanthus* [34]. In this model, the space is 3-D hexagonal lattice and cells are modeled as elongated ellipsoids. It was assumed that C-signaling occurs when cell poles of neighboring cells overlap and that the cells reduce their speed in a traffic jam. This cellular automata model qualitatively describes several stages of the aggregation process. The traffic jam causes the initial aggregation of cells. Cells thereafter form streams in moving towards the aggregates and some cells circulate the aggregates. Following this work, Holmes et al developed another 3-D cellular automata model for the intercellular interaction during aggregation into fruiting bodies. This model reproduced several multicellular behaviors, such as aggregation center formation, aggregate formation and dispersion[41].

The cellular automata model is commonly used in mathematics, physics and theoretical biology[96-99]. These models discretize the space into grids, and therefore, lack the scalability going from a very small dimension inside a cell body to an extremely large dimension that may contain hundreds of thousands of cells. Also, the models lack flexibility in the definition of cellular behaviors, intracellular processing and intercellular variation. Agent-based modeling is a far superior alternative for modeling individuals with volition such as microorganisms.

1.2.3. Agent-based models

The agent-based model (ABM) is a powerful tool in modeling socially interacting agents [100, 101]. In ABM, the functions and the modular implementations of various components of behavior for individual agents can be

easily modeled. Moreover, the ABM allows to model social interactions between individuals. Agent-based modeling can be used to solve numerous business and technology problems such as supply chain optimization and logistics [102], modeling of consumer behavior [103], workforce management [104], and traffic analysis [105]. Applications of ABM in biology include modeling population dynamics [106], ecological systems [107], inflammation [108], cancer and tumor [109], wound healing [110], vascular biology [111], intercellular signaling [112], infectious diseases [113], as well as human immune systems [100, 114].

There are two crucial steps in agent-based models: validation and verification. The validation step is to verify that all implemented methods are mathematically correct and all numerical errors are excluded. The verification part is to confirm that the results of agent-based simulations agree with experimental observations. Agent-based model simulation can capture both transient and steady states of a system. A successful ABM reproduces the natural behavior of individuals as well as the collective patterns of groups of individuals.

The properties of ABM make it suitable for modeling groups of individual cells. In 2005 Anderson et al developed a 1-D agent-based model of developmental rippling following the work of Igoshin et al. in 2001[7, 62]. In 2006, Sliusarenko et al incorporated the *Frz* model and head-to-head C-signal transduction into their agent based model and successfully reproduced experimental observations of ripples [36]. and, further, fruiting-body aggregation. The authors obtained qualitative agreement with experimental observations and pointed out that the traffic jam in a high cell

density area could be the mechanism behind the *M. xanthus* aggregating into fruiting bodies[39]. In the meantime, Wu et al studied the swarming in *M. xanthus* using an individual based model. The authors demonstrated quantitative agreement on dependencies between swarm expansion rate and cell density in the colony. One of the findings of Wu et al was that the natural reversal period of around 8 minutes for individual cells was not a coincidence; it was actually the best reversal time to achieve the largest swarm expansion rate[40, 95].

1.2.3.1. Agent-based modeling of swarm expansion

Studying the motility mutations can yield insights into the mechanisms behind the swarm expansion patterns. Experimental observations show that wild type swarms have the largest expansion rate. A-motile swarms have the second largest expansion rate and S-motile swarms have the smallest expansion rate under the same conditions. Moreover, when the cell density is low, the expansion rates for all three strains show a positive correlation with cell density. However, when the initial cell densities reach thresholds, the expansion rates for all three kinds of strains saturate [66]. These are the essential characteristics of the swarming pattern.

Wu et al developed an individual based model to study the expansion phenomenon [40, 95]. In their model, each cell is described as a system with three nodes connected by two segments. The expansion rate was modeled as the cell flux into the simulation domain. This assumption was based on the following arguments.. First, the expansion rate in the agent-based simulation was difficult to

measure since only a small region close to the leading edge of the colony was simulated. Second, it has been observed that at steady state, , *M. xanthus* cells always form a one-cell layer, full-coverage colony in the region close to the leading edge of the swarm. Therefore, the cell density should be constant in this region. The authors compared the expansion rates between A-motile and wild type strains with various initial cell densities and achieved qualitative agreement in the swarm expansion: the expansion rate first increases then saturates at a certain cell density and the wild type swarm expands faster than the A-motile strain [66]. Later on, they claimed that the 8 minutes of reversal period found in real *M. xanthus* cells are optimized for fastest swarming using the same model. However, the authors did not test the S-motile strains' behavior although S-motility was implemented in the agent-based model.

1.2.3.2. Agent-based modeling of rippling patterns

Sliusarenko et al developed an agent-based model of *M. xanthus* rippling [36]. The model has two assumptions. One assumption is that individual cells move back and forth controlled by an internal biochemical clock. The other assumption is that head-to-head contact mediated signaling can change the cellular reversal periods. Newton's laws of motion were implemented to compute the trajectories and *Frz* model proposed by Igoshin et al [56] was adopted for modeling the cellular reversals. The agent-based model reproduced the spatial and temporal synchronization of motions of *M. xanthus* cells as observed in their experimental work. The model also reached a quantitative agreement of reversal period

distributions between simulations and experiments for both before and after rippling. Moreover, the model revealed that wavelengths of ripples are about twice the distance cells travel before they reverse. Finally the model showed that the ripple wavelength is positively correlated with the fraction of the C-signal defective cells in the total *M. xanthus* population. This observation is verified by Sliusarenko et al's experimental data. Using the agent-based model, the authors also revealed that when two opposing wave crests collide, the cells in the leading edge of each crest reverse first and cause the following cells to reverse.

The agent-based model made the dynamics of developmental ripple clear. However, recent observations by Mauriello et al indicated that there could be alternative signal mechanisms that lead cells to form a ripple patterns during *M. xanthus* predation [59]. The model needs to be reconstructed to fit the new observations.

1.2.3.3. Agent-based modeling of aggregation into fruiting bodies

Under harsh environmental conditions, *M. xanthus* swarms show coordinated motions of aggregation into fruiting bodies. The whole developmental process usually takes less than 24 hours. The agent-based model developed by Sliusarenko et al has successfully uncovered sets of intercellular interactions that lead to patterns similar to those observed experimentally in *M. xanthus*[39]. The model is based on the traffic jam principle: cells tend to slow down or stop when entering regions of high cell density. The traffic jam leads to a positive-feedback loop as more cells jam into the region, further increasing cell density so that aggregates are able

to grow without any diffusible morphogens or chemoattractants. Sliusarenko et al discovered that traffic jamming is capable of producing aggregates of *M. xanthus* cells. Furthermore they studied the effect of intercellular interactions on the formation of aggregates. They found out that changes of C-signal and alignment affect the final size and shape of the aggregates. However, their research lacks comparison of the dynamics of the aggregation process. There is no quantitative match for the aggregate size and number between simulations and experiments in their study. Moreover, the authors focused on the formation of single aggregates rather than properties associated with multiple aggregates. The research of Sliusarenko et al does not provide enough evidence for the changing spatial order of aggregates of *M. xanthus* cells.

1.3. Summary and motivation

In recent years the ubiquity of microbial communities in nature has become apparent: for instance, most bacteria related to human diseases are associated with biofilms. However, it is still not clear how individual cells self-organize in these communities and how community as a whole responds to environmental cues. The work presented in this thesis aims to discover the mechanisms of self-organization in dynamic single-species biofilms (swarms). In these swarms, bacteria display collective surface motility, cooperatively sense the environment, execute collective developmental programs and often differentiate into distinct cell types performing specialized functions. Research on mechanisms of biofilm formation addresses questions similar to those in developmental biology: how to connect macroscopic

phenotypes and biochemical pathways in individual cells? Researchers know the genome composition of the cells and can reveal the mutations that affect emergent spatio-temporal patterns, but it remains a challenge to relate these mutations to mechanisms of intercellular signaling and motility. Decoding these mechanisms from phenotype observations is a complex reverse-engineering problem that cannot be solved solely by traditional experimental research. In this work complementary approach is evoked a combination of agent-based modeling and biostatistical image quantification. The research is focused on self-organization in spreading or aggregating biofilms formed by *Myxococcus xanthus* – model organism to study microbial cooperation, development and collective motility. This bacterium uses two motility systems and multiple sensory and signaling pathways to move over surfaces forming varieties of population patterns. These patterns reveal motility coordination of individual cells as well as their ability collectively sense and respond to environmental cues. In this thesis, I present my work towards understanding the dynamics in the aggregation into fruiting bodies and the mechanisms and physiological role of predatory rippling.

Chapter 2

Methods

2.1. Agent-based models

Agent-based modeling is a widely used approach to modeling systems with interacting autonomous agents, and is a powerful tool to study how individual cell activities and environment affect the organization of a system[36, 39, 40, 62, 72, 95]. Agent-based models can incorporate cell variability and achieve modular implementation of intercellular interactions. In this thesis, I will introduce two agent-based models that I developed for *M. xanthus* fruiting body aggregation and predatory ripple formation.

2.1.1. Agent-based *in silico* model of *M. xanthus* development

To simulate traffic-jam aggregation, this agent-based model follows the model developed by Sliusarenko et al [39]. Each agent represents a cell characterized by its length *(L)*, width *(W)*, center position *(x,y)*, orientation

$\theta(0 \le \theta \le 2\pi)$ and speed (v). The physical parameters L and W are constant throughout the simulations, while other parameters are updated at each time step. In addition, to model the regular reversals of agents, the reversal is mapped to a circle by introducing a new parameter phase φ $(0 \le \varphi \le 2\pi)$ (Figure 2-1). The phase models the reversal clock of the cell [56], with phase points 0, π and 2π corresponding to changing of the direction polarity. In the agent-based model, the phase advance speed is based on the reversal period T:

$$\omega = \frac{\pi}{T} \qquad \text{Eq. (1)}$$

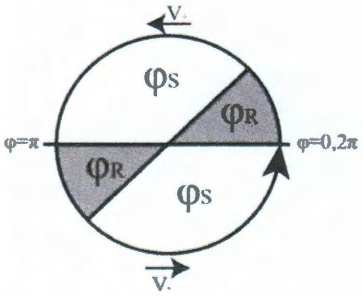


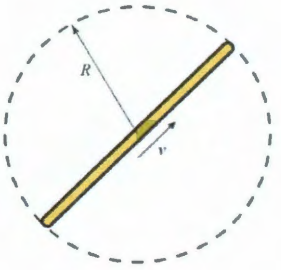
Figure 2-1 Phase parameter is modeled as a reversal clock of a cell

For each simulation, the model has a time step (δt), space dimensions $(X$ and $Y)$, and initial agent density (ρ). The number of agents (N) that are needed for each simulation can be calculated by $N = \rho XY$.

During a simulation, periodic boundary conditions are applied so that all agents are bounded inside the space domain. Agents are updated sequentially during each time step in a simulation as follows: at time t, denote the ith updated agent as having the center position $(x_i(t), y_i(t))$ with orientation $\theta_i(t)$ moving direction $\alpha_i(t)$ in phase $\varphi_i(t)$ at speed of $v_i(t)$. The number of agents $(n_i(t))$ in the

round area around $(x_i(t),y_i(t))$ with radius R (Figure 2-2) is counted, and the local agent density in the head area is computed as $\rho_{i,local} = \frac{n_i(t)}{\pi R^2}$.

Figure 2-2 Definition of a neighborhood area of an agent



2.1.1.1. Phase update

The phase $\varphi_i(t)$is updated so the phase for one time step forward $\varphi(t+\delta t)$ is: $\varphi_i(t+\delta t) = \varphi_i(t) + (\omega + r_\varphi(t)) \cdot \delta t$, where $r_\varphi(t)$ is the random noise term: $r_\varphi(t+\delta t) = r_\varphi(t) + \sqrt{6D_\varphi \delta t} \cdot U(-1,1)$. $U(a,b)$ denotes a random number generated by uniform distribution in the region (a,b), and D_φ is the phase diffusion coefficient that characterizes random fluctuations in phase. Phase advance might be negative because of random fluctuations, but phase is not allowed inversely to pass 0, π:

$$\varphi_i(t+\delta t) = \begin{cases} 0, & \text{if } \varphi_i(t) > 0 \text{ and } \varphi_i(t+\delta t) < 0 \\ \pi, & \text{if } \varphi_i(t) > \pi \text{ and } \varphi_i(t+\delta t) < \pi \end{cases}$$

and $\varphi_i(t+\delta t)$ is assured to be bounded by 2π as:

$$\varphi_i(t+\delta t) = \varphi_i(t+\delta t) - 2\pi, \text{ if } \varphi_i(t+\delta t) > 2\pi.$$

C-signaling occurs between leading poles of two adjacent cells. The coordinates of leading poles are calculated by:

$$x_{i,head}(t) = x_i(t) + \frac{L}{2}\cos(\theta_i(t)),\ y_{i,head}(t) = y_i(t) + \frac{L}{2}\sin(\theta_i(t)).$$

A distance $d_{h,i,j}(t)$ is defined between ith cell leading pole and jth cell leading pole:

$$d_{h,i,j}(t) = \sqrt{(x_{i,head}(t) - x_{j,head}(t))^2 + (y_{i,head}(t) - y_{j,head}(t))^2}\ .$$

The C-signaling is implemented so that all leading poles within a distance of R are checked pairwise for exchanging signals, and the phase is updated by a phase resetting function $\Psi(\varphi)$ multiplied by α, the signaling strength . The streaming signaling is implemented so that the phase advance speed ω decreases with inceasing cell density (ρ_1 is the density dependent factor). As a result, the phase advance considering the C-signaling and streaming signaling is modeled as:

$$\varphi(t+\delta t) = \varphi(t) + \left(\frac{\omega}{1+\rho_{i,local}/\rho_1} + \alpha \frac{\sum_{j:d_{h,i,j}<R_c} \cos^2[(\theta_i(t) - \theta_j(t)/2]}{\pi R_c^2} \psi(\varphi) + r_\varphi(t) \right) \delta t.$$

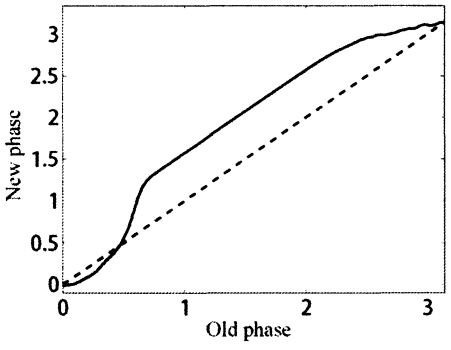


Figure 2-3 Phase resetting function $\Psi(\varphi)$ in $(0,\pi)$ (solid line). Immediately after reversal, the head-head signaling decreases the phase. After a time period, the signal leads to an increase in the phase.

2.1.1.2. Speed update

The speed of the agent is then updated with respect to the local cell density $\rho_{i,local}$ and the density threshold ρ_0. v_0 is the normal cell speed. τ_v is the speed correlation time. v_{low} is the reduced speed at high cell density. $H(x)$ is a Heaviside step function. D_v is the diffusion coefficient that characterizes random fluctuations in speed. $v_i(t+\delta t)$ is calculated by solving the ODEs:

$$\frac{dv}{dt} = \frac{v_0}{\tau_v} - \frac{v}{\tau_v} + r_v(t), \quad \rho_{i,local} < \rho_0$$
$$\frac{dv}{dt} = \frac{v_0}{\tau_v} - \frac{v_0 v}{v_{low}\tau_v} + r_v(t), \quad \rho_{i,local} < \rho_0$$

using the implicit Euler method finite-difference scheme as follows:

$$v_i(t+\delta t) = \begin{cases} \dfrac{v_i(t) + (\frac{v_0}{\tau_v} + r_v(t))\delta t}{1 + \frac{\delta t}{\tau_v}}, & \rho_{i,local} < \rho_0 \\ \dfrac{v_i(t) + (\frac{v_0}{\tau_v} + r_v(t))\delta t}{1 + \frac{v_0 \delta t}{v_{low}\tau_v}}, & \rho_{i,local} > \rho_0 \end{cases}$$

The fluctuation in velocity due to random force is given by the following random term: $r_v(t) = \sqrt{6D_v \delta t}U(-1,1)$.

2.1.1.3. Orientation update

Orientation $\theta_i(t)$ is updated according to the alignment between cells, cell density gradients and phase: $\theta_i(t+\delta t)=\theta_i(t)+\Omega_i(t)\delta t+\pi G(\varphi_i(t))$. The $G(\varphi_i(t))$ is defined to increase the orientation by π when phase $\varphi_i(t)$ crosses 0, π, and 2π:

$$G(\varphi_i(t+\delta t))=\begin{cases}1, & \{\varphi_i(t)<\pi \text{ and } \varphi_i(t+\delta t)>\pi\} \text{ or } \{\varphi_i(t)<2\pi \text{ and } \varphi_i(t+\delta t)>0\}\\ 0, & \text{otherwise}\end{cases}$$

In this model, each agent aligns with other agents in the round neighborhood with radius *R*. At the same time, they also align to the cell density gradient when the gradient is sufficiently large. τ_θ is the relaxation time of angular velocity. $\tau_{\theta 1}$ is the relaxation time for cells alignment with other cells. $\tau_{\theta 2}$ is the relaxation time for cells to align with density gradients. $\theta_{g(\rho)}$ is the direction from the center of the *i*th cell to the mass center of all cells in the neighboring area of the *i*th cell.

$$\Omega_i(t+\delta t)=\Omega_i(t)+\{-\frac{\Omega_i(t)}{\tau_\theta}-\sum_{j\ \text{in}\ n_i(t)}\frac{\sin[2(\theta_i(t)-\theta_j(t)]}{\tau_{\theta 1}}+\frac{\sin\{2[\theta_i(t)-\theta_{g(\rho)}]\}}{\tau_{\theta 2}}\cdot H[g(\rho)-g_0]+r_\theta(t)\}\delta t$$

The random fluctuation in angular velocity is characterized by D_θ, as:

$r_\theta(t)=\sqrt{6D_\theta\delta t}U(-1,1)$.

2.1.1.4. Position update

Finally, the *i*th agent is moved to a new position defined by its current speed and orientation: $x_i(t+\delta t)=x_i(t)+v_i(t)\cdot\delta t\cdot\cos(\theta_i(t))$, $y_i(t+\delta t)=y_i(t)+v_i(t)\cdot\delta t\cdot\sin(\theta_i(t))$.

Table 1 Parameters used in the ABM for *M. xanthus* development

Parameter	Value	Description
L	5μm	Agent length
W	0.5μm	Agent width
R	5μm	Radius of interaction area (for calculating local cell density)
ρ	0.3 cell/μm^2	Initial agent density
δt	3s	Time step
X x Y	1000x1000μm^2	Simulation domain
V	0.1μm/s	Normal agent speed
v_{low}	0.02μm/s	Low agent speed (traffic jam)
ρ_0	0.7 cell/μm^2	Agent density threshold
ω	0.008rad/s	Phase speed
$\tau_{\theta 1}$	10000s	Relaxation time for aligning to neighboring cells
$\tau_{\theta 2}$	10000s	Relaxation time for aligning to density gradients
τ_v	50s	Relaxation time for cell speed
τ_θ	50s	Relaxation time for angular velocity
D_v	0.0001μm^2/s	Spatial diffusion coefficient
D_θ	0.0001rad^2/s	Direction diffusion coefficient
D_φ	10^{-4}rad^2/s	Phase diffusion coefficient
g_0	0.3/μm^2	Density gradient threshold for alignment
ρ_1	17/μm^2	Density regulation on reversal periods
A	1/μm^2	Head-to-head signaling coefficient
R_c	1μm	Signal interaction distance

2.1.2. Agent-based models for *M. xanthus* ripple pattern

This model is an extension of the earlier ABM of *M. xanthus* self-organization [36] with incorporation of the new side-to-side signaling mechanism. Similar to the agent-based model of aggregation, each agent represents a cell, a self-propelled rod on a 2-D surface with length of L, width of w, with a center position of $(x(t),y(t))$, and orientation $0 \le \theta(t) \le 2\pi$. Agent length and width are also kept constant throughout all simulations whereas the center position and moving direction are changed each time step as cells move and align. For each simulation, time advances with constant increments δt. The simulations are conducted on a fixed 2-D area in which all simulated moving agents are bounded. For most simulations I impose periodical boundary conditions.

2.1.2.1. Cell movement

Agents' center positions are updated at each time step with both directed and random displacement:

$$\begin{aligned} x(t+\delta t) &= x(t) + v \cdot \delta t \cdot \cos(\theta) + \sqrt{D\delta t} \cdot U(-1,1) \\ y(t+\delta t) &= y(t) + v \cdot \delta t \cdot \sin(\theta) + \sqrt{D\delta t} \cdot U(-1,1) \end{aligned}.$$

Here v is the average cell speed whereas D is the effective diffusion coefficient corresponding to fluctuations of speed. These parameters are estimated from experimental data. Here and below $U(a,b)$ denotes a random number generated by a uniform distribution between a and b.

2.1.2.2. Cell reversals

In order to track timing between cell reversals, an internal timer phase variable, $\varphi(t)$, is introduced which is ranged in $[0,2\pi)$. At each time step, the phase advances and when the phase increases past π and 2π, agents change their orientation θ by 180 degrees. Without any signaling, the reversal period is T and so the average phase speed is given by Eq. (1).

As a result, without signaling, each time-step the phase is updated as follows:

$$\varphi(t+\delta t)=\varphi(t)+\omega\cdot\delta t+\sqrt{D_{\varphi}\delta t}U(-1,1)$$

where D_{φ} is the effective diffusion coefficient of phase, characterizing the fluctuation in phase velocity or equivalently fluctuations in reversal time. The value of term D_{φ} is obtained by matching the reversal period distributions of simulation and experimental observations in section 2.3.4. After $\varphi(t)$ is computed at each time step, the following procedures are applied to periodically bound $\varphi(t)$ within $[0,2\pi)$ and ensure that the random fluctuations in phase near π and 2π do not lead to additional reversals. The random fluctuations near π and 2π are dealt with first

$$\varphi(t+\delta t)=\begin{cases}0, & \text{if } \varphi(t+\delta t)<0\\ \pi, & \text{if } \varphi(t+\delta t)<\pi \text{ and } \varphi(t)>\pi\end{cases}.$$

Then the periodic boundary condition is applied:

$$\varphi(t+\delta t)=\varphi(t+\delta t)-2\pi, \text{ if } \varphi(t+\delta t)\geq 2\pi.$$

When the phase increasing crosses π or 2π, the cells reverse its direction by switching the polarities of its two ends:

$$\theta(t+\delta t) = (\theta(t)+\pi) \bmod (2\pi)$$
$$\text{if } (\varphi(t) < 2\pi \text{ and } \varphi(t+\delta t) > 0) \text{ or } (\varphi(t) < \pi \text{ and } \varphi(t+\delta t) > \pi) \qquad \text{Eq. (2)}$$

2.1.2.3. Side-to-side signaling mechanism and induced reversals

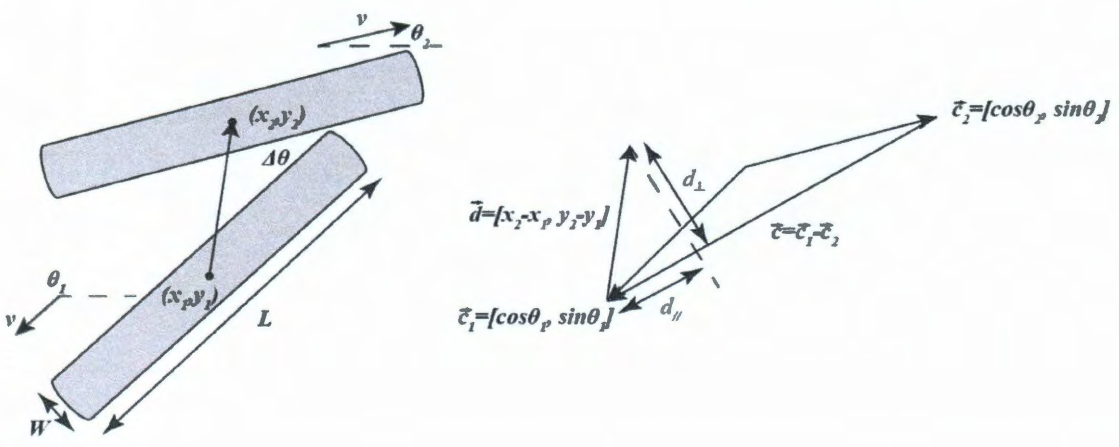


Figure 2-4 Side-to-side signaling in ABM. The side-to-side contact in the ABM is defined by 3 parameters: The perpendicular (to cell orientation) distance between the center of the two agents ($d_\perp$), the parallel distance between the center of the two agents ($d_{//}$) and the angle formed by the two agents (Δθ in this figure). L represents the length of the cells and v represents velocity.

Besides the regular cell reversals, signal induced early cell reversals are the key for the ripple pattern. Recent experimental results show that during the side-to-side contact of two *M. xanthus* cells, their *FrzCD* clusters align and as a result one of the cells could reverse [59]. Based on the experimental observation, an assumption is made that the observed side-to-side signaling is able to induce early cell reversals. An algorithm is then developed that implements the side-to-side signal mechanism in this agent-based model. The algorithm obtains the spatial relationships between

neighboring agents and checks if the relative positions of two neighboring cells permit a side-to-side contact signaling. The algorithm also describes the way an agent respond to the signal. For a selected agent, the algorithm first generates a list of neighboring agents whose centers are inside a local square region centering around the selected agent center. Then the following procedures are applied pairewisely between the selected agent and one of its neighbors to determine if these two agents satisfy the conditions for side-to-side signaling and how to react if the signal happens. The side-to-side signaling only happens between two touching agents with their long axis aligned and traveling in opposite directions (see Figure 2-4). Therefore the following conditions are imposed to detect agents that made side-to-side contact:

(1) To detect two agents with orientations θ_1 and θ_2 that have nearly parallel long axis but travel in opposite directions the algorithm seeks cells for which:

$$\left| \left| \theta_1 - \theta_2 \right| - \pi \right| < \Delta\theta_0. \qquad \text{Eq. (3)}$$

For the simulations preformed the threshold is $\Delta\theta_0$ to be 0.083π=15 deg.

(2) To check if two nearly parallel agents touch and have significant overlap along their long axis the algorithm places two additional conditions on the proximity of cells. Suppose (x_1,y_1) denotes the center of the selected agent and (x_2, y_2) denotes one of its neighbors. A vector d is defined from the center of the selected agent to its neighbor: $d = [x_2 - x_1, y_2 - y_1]$. For two cells to touch limits are set on the projection of the vector d on the axis along the cell length and in the

perpendicular direction (see Figure 2-4). To define the average direction of the two cells, unit vector $c_i = [\cos(\theta_i), \sin(\theta_i)]$ is used to represent orientation of *i*th represent agents. As seen in Figure 2-4, vectors c_1 and c_2 would be pointing in nearly opposite direction therefore the average orientation of two cells is along the axis determined by a vector $c = c_1 - c_2$

The vector d then can be projected into the average cell orientation defining the separation of cell centers along their average direction $d_{//}$ and onto perpendicular direction (distance $d_{\perp}$) as follows (Figure 2-4): $d_{//} = | d \cdot \frac{c}{||c||} |$,

$d_{\perp} = || d - d_{//} \frac{c}{||c||} ||$, where ||...|| is the vector norm and |...| is an absolute value. For parallel two cells to make side-to-side contact their $d_{\perp}$ must not exceed cell width:

$$d_{\perp} \leq w \qquad \text{Eq. (4)}$$

At the same time, for efficient signaling cells must overlap at least 50% of their long axis length. It results in

$$d_{//} \leq \frac{L}{2} \qquad \text{Eq. (5)}$$

Conditions Eq. (3), Eq. (4) and Eq. (5) are checked every time steps to determine all cell pairs that are in side-to-side contact and therefore capable of signaling. However, It is assumed that not every side-to-side contact results in signaling event and therefore introduce a parameter, p_0 probability of signaling

given the side-to-side contact. It is assumed that the signaling is asymmetric and that probabilities of cell #1 to signal to cell #2 and vice versa are statistically independent. It is also assumed that $p_0 << 1$. These assumptions are motivated by the observations of Mauriello et al [42]that only of one of two cells reverses as a result of side-to-side contact. Therefore, for each cell in a side-by-side contact pair the algorithm generates a random number $U(0,1)$ and only consider signaling to occur if $U(0,1) < p_0$.

Every successive signaling event will result in the reversal unless cell is refractory, i.e. have recently reversed. As each reversal event is associated with a flip of cellular polarity and requires reorganization of cellar motors it is natural to assume that there is a minimal reversal period during which cell will not be able to reverse again. This period is called a refractory period and measure it using a phase variable clock. After each reversal there is a sector φ_0 in phase clock corresponding to an average refractory time T_0, and $\varphi_0 = \omega T_0$. During this time, an agent does not respond to the side-to-side signal but it can always actively signal to other agents. Otherwise, if the agent is not refractory, i.e. if: $\varphi_0 < \varphi(t) < \pi$ or $\varphi_0 + \pi < \varphi(t) < 2\pi$, the agent is responsive to signal and reverse with the probability p_0. After the reversal, the phase variable is reset as: $\varphi(t+\delta t) = \begin{cases} 0, & \varphi(t) > \pi \\ \pi, & \varphi(t) < \pi \end{cases}$. As the signal can induce agent reversal with the probability p_0, the orientation of the agent is reset upon reversal as:

$$\theta(t+\delta t) = (\theta(t) + \pi) \bmod (2\pi) \qquad \text{Eq. (6)}$$

2.1.2.4. Cell alignment

Local alignment of cells is essential for rippling. In this model, cells are modeled as rods that align to each other according to the equations introduced by Sliusarenko et al [36]. More sophisticated alignment algorithms are not feasible in this agent-based model, as simulations with up to 300,000 cells are required.

$$\frac{d\theta_i}{dt} = \frac{\sin(2(\theta_i(t) - \theta_i^0(t)))}{\tau_\theta} + r_\theta(t) \qquad \text{Eq. (7)}$$

where τ_θ is the angle correlation time and $r_\theta(t)$ is the random noise. θ_i^0 is the average nematic orientation of neighbors computed as follows. First a neighboring region is defined around each agent. To ensure fast computational speed, a square region is used with dimensions centered in the center of the selected agent. At each time step, for each agent i, a list of n_i neighbors is found with centers inside the square region. The algorithm then computes θ_i^0 the average orientation of neighbors as follows:

$$2\theta_i^0 = \arctan\left(\frac{\sum_{j=1}^{n_i} \sin(2\theta_j)}{\sum_{j=1}^{n_i} \cos(2\theta_j)}\right).$$

The algorithm then discretizes Eq. (7) using implicit finite difference scheme to be solved iteratively

$$\theta_i(t+\delta t) = \theta_i(t) - \frac{\sin(2(\frac{\theta_i(t)+\theta_i(t+\delta t)}{2} - \theta_i^0(t)))\delta t}{\tau_\theta} + \sqrt{D_\theta \delta t} U(-1,1)$$

If no reversal happens, the computed $\theta_i(t+\delta t)$ is the orientation of agent *i* at the next time step. Otherwise, Eq. (2)and Eq. (6) are used to further update the orientation.

2.1.3. Parameter selection and optimization

For the agent-based model of *M. xanthus* development discussed in section 2.1.1, some parameters can be collected from experimental observation, such as cell dimension, speed, and reversal period, but many parameters must be based on best estimations and assumptions. When an agent-based model is used to simulate a biological system, relatively small changes in parameter values may alter the emerging pattern dramatically. Therefore, to achieve a pattern in the simulation that closely matches the experimental observations, a parameter optimization algorithm is needed. When possible, the agent-based model parameters were adapted from the previous work of Sliusarenko et al [39] as discussed, but nine parameters had to be tuned using the optimization algorithm to achieve agreement with four experimentally observed aggregation features that characterize the dynamics during aggregation and steady state aggregation pattern: (i) final aggregate number, (ii) final aggregate size, (iii) aggregate growth rate and (iv) aggregate shape. The parameter optimization algorithm employs a genetic algorithm of real optimization with modifications that meet the requirement of this work. The parameter combinations of the first generation were selected to produce aggregates that have

properties comparable to experimental observations (i) – (iv). Parameters corresponding to Sliusarenko et al [39] were selected as starting values. Each subsequent generation had 60 mutated parameter sets and thus 60 new simulations with these parameters were initiated. After the simulations finished, the results were then fed into a fitness function that computed agreement between simulation results and experimental observations. Ten parameter combinations with the best fitness were used to generate 60 new sets using the same mutation method. After approximately 10 generations, little improvement in the fitness function was observed, and the parameter combination with best fitness was used in this agent-based model. This optimization was geared towards finding a local optimum, and further improvement could theoretically be possible in different regions of the parameter space. The detailed genetic algorithm is described as following.

The algorithm starts with a parameter set used by Sliusarenko et al. [39] and perturbs these parameters to generate 10 initial distributions of parameters. After the first generation parameter sets are chosen, they are fed into new agent-based model simulations. Each simulation outputs the agent density every 2 minutes, and the cell density results are used to compute fitness. For each generation, 60 new parameter combinations are created by mutating the 10 best-fit parameter combinations of the previous generation. Suppose that $h_k(g)$ is the k^{th} parameter in one parameter combination at generation g, σ_k is the constant step size for the k^{th} parameter, and (h_{ka}, h_{kb}) is the range of the parameter, the parameters are then mutated at the rate p_m for next step:

$$h_k(g+1)=\begin{cases} U(h_k(g),\min(h_k(g)+\sigma_k,h_{kb})), & p<p_m \text{ and } q<0.5 \\ U(\max(h_k(g)-\sigma_k,h_{ka}),h_k(g)), & p<p_m \text{ and } q>0.5 \\ h_k(g), & \text{otherwise} \end{cases}$$

where p, $q=U(0,1)$ are random numbers. Table 2 shows one starting parameter combination with its parameter range and step sizes.

Agent densities are analyzed at the end of each simulation by the same technique used to identify aggregates in experimental images (see later in 2.2.1.1). Four aggregate measurements are computed as a fitness function: the number of fruiting bodies in a 1mm^2 area, the average size of fruiting bodies in μm^2, the growth rate in μm^2/min, and the shape penalty (eccentricity e, described below). Every measure of fitness follows a weighted probability density function of Gaussian distribution with a cutoff at 0 and maximum value of 1. The mean and standard deviation of the weighted Gaussian probability density function are estimated from experimental observations. For a particular measure, the fitness (F_υ) can be calculated by the value υ computed from simulation results with mean υ and standard deviation σ estimated from experimental observations:

$$F_\upsilon = \exp(-\frac{(\upsilon-\bar{\upsilon})^2}{2\sigma^2}).$$

The sum of the four measures is the total fitness:

$$F = F_{num} + F_{size} + F_{growthrate} + F_{shape}.$$

Thus the possible maximum fitness is 4. The closer the fitness is to 4, the better the parameter combination.

The observed aggregate number has a mean of 9.1 and standard deviation of 1.8. So the fitness function of aggregate number is:

$$F_{num} = \exp(-\frac{(n-9.1)^2}{2\times 1.8^2}),$$

where n is number of aggregates in 1 mm^2. Similarly, the experimentally observed average aggregate size is 1.66×10^4 μm^2, with standard deviation 6.4×10^3 μm^2, so the size fitness function is:

$$F_{size} = \exp(-\frac{(s-1.66\times10^4)^2}{2\times(6.4\times10^3)^2}),$$

where s is the average size of aggregates in simulation results. Since the growth of aggregates is not monotonic, the experimental growth rate is estimated in the rapidly increasing phase. The average growth rate is 27.13 μm^2/min with standard deviation 17.63 μm^2/min, so the fitness function is:

$$F_{growthrate} = \exp(-\frac{(\lambda-27.13)^2}{2\times 17.63^2}),$$

where λ is the growth rate in simulation results. The two dimensional shape of a fruiting body is assumed to be close to an ellipse. The eccentricity parameter is used to characterize the shape of an aggregate. The eccentricity (e) of a random shape is defined as the eccentricity of an ellipse that has the same second moment as a

random shape. The eccentricity of an ellipse is the ratio of distance between two focal points to the length of the major axis. Therefore, e is in [0,1]. When the shape is round, e=0; when the shape is a line segment, e=1. The experimentally measured mean eccentricity of fruiting bodies is 0.3 and the standard deviation is 0.17. In the parameter optimization, the fitness function of aggregate shape is:

$$F_{shape} = \exp(-\frac{(e-0.3)^2}{2\times 0.17^2}).$$

Table 2 Parameter range used in optimization

Parameter	Units	Starting value	Range	Step
ρ_0	cell/μm²	0.7	0.5-2.0	0.1
v_{low}	μm/s	0.02	0.0001-0.05	0.0002
$\ln(D_v)$	ln(μm²/s)	-12.72	(-6.92)-(-18.42)	1
$\ln(\tau_{\theta 1})$	ln(s)	6.91	4.6-9.2	0.3
$\ln(\tau_{\theta 2})$	ln(s)	6.91	4.6-9.2	0.3
$\ln(D_\theta)$	ln(rad²/s)	-12.72	(-6.92)-(-18.42)	1
$\ln(D_\varphi)$	ln(rad²/s)	-9.21	(-2.3)-(-18.42)	1
P_1	μm^{-2}	2.5	1-20	1
A	μm^{-2}	5	0-20	1

The parameters used for agent-based simulations for predatory ripple pattern (2.1.2) are summarized in Table 3. Whenever possible parameters used were estimated directly or indirectly from the experimental data obtained. For instance, the analysis of individual cell movement described above not only

provides the average behaviors of cells such as average velocity and reversal period but allows with their population distributions. The agent velocity v used in ABM simulation is the average velocity calculated in the above analysis. The diffusion coefficient D, which is characterizing the random fluctuation in agent movement, is chosen so that the variance of the instant velocity distribution of ABM simulation matches the results of experimental data analysis. Note that experimental observed random fluctuations along x direction and y direction are almost identical. Only one value D is used to represent the noise level in cell movement. In the ABM simulations of rippling, varying the refractory period changes the average reversal period. Thus the refractory period is chosen in a way that can lead to the same average reversal period in ABM as in experimental observations of rippling cells. The average reversal period of non-rippling cells observed experimentally is chosen as the natural reversal period T in the ABM and the phase speed ω is calculated using Eq. (1). The diffusion coefficients in reversal period D_φ is chosen by matching the distribution of reversal periods of ABM simulations to the experimental observed distribution. The phase variable φ_0 in ABM simulations are chosen in a way that φ_0 / ω should equal the selected refractory period. There are also parameters cannot be estimated experimentally directly but can be limited based on the simulation results. For example, the random noise level D_θ is chosen such that the initially aligned population of cells remains aligned.

Table 3 Parameter used in agent-based model of *M. xanthus* ripple pattern

Variable name	Description	Value	Source
P	Cell density	0.3 /μm²	The cell density is corresponding to one layer of densely packed cells
Δt	Simulation time step	5 s	
W	Cell width	0.5 μm	From observation of individual cells
L	Cell length	7 μm	From observation of individual cells
V	Agent velocity	6 μm/min	Average cell velocity observed in experiments
D	Spatial diffusion coefficient	0.1 μm²/s	Obtained by matching instant cell velocity distributions in ABM simulations and experimental data analysis
T	Natural reversal period of agents	8 min	Average cell reversal period observed in experiments of non-rippling cells
D_φ	Phase diffusion coefficient	10^{-4} rad²/s	Obtained by matching reversal period distributions in ABM simulations and experimental data analysis
T_0	refractory time	2.6 min	Obtained by matching the average reversal period in ABM simulations and experimental observations of ripping cells.
$\Delta\theta_0$	Angle difference threshold for parallel agents	0.0833π	
τ_θ	angle correlation time	50 s	Adapted from the work of Sliusarenko et al.
D_θ	Angle diffusion coefficient	10^{-4} rad²/s	Adapted from the work of Sliusarenko et al.
p_0	Signal probability	0.1	Toned to match the experimental observed pattern

2.1.4. Implementations of the agent-based models

The agent-based models for aggregation and rippling were implemented in Java programming language with Java SE 1.5. The implementation takes advantage of the Repast agent simulation toolkit [115] as well as Java concurrency [116]. Individual simulations usually take twenty to one hundred CPU hours on STIC (Shared Tightly Integrated Cluster) and SUG@R (Shared University Grid at Rice) at Rice University [117] depending on the size of the simulations. Number of agents, dimensions of a simulation domain, the distribution of agents and the complexity of intercellular interactions together determine the size of a simulation.

The steps in an agent-based model are described in Figure 2-5. To speed up the process, two techniques are employed. Multithreading is implemented that each thread processes a sub-list of agents based on the independence between agents. Thus, the CPU hours required decrease 2-8 fold comparing to single thread, sequential implementations on STIC and SUG@R. To efficiently obtain local neighbor list, the simulation domain is partitioned into small subspaces. A hash-table is created for each subspace to maintain the list of agents centered in the subspace at current time. When searching for a list of neighbors of an objective agent, the program only searches for agents in subspaces near the objective agent.

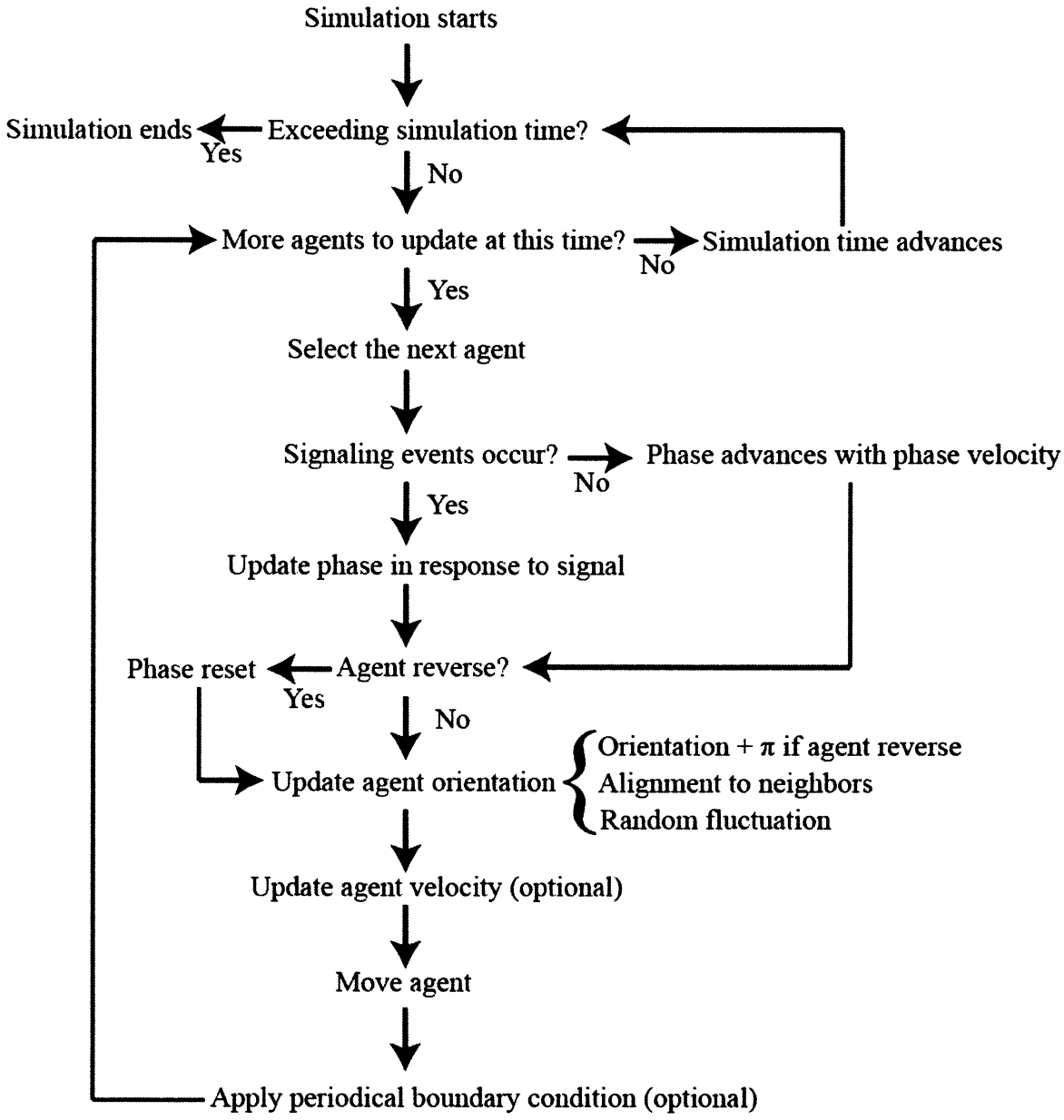


Figure 2-5 Flow chart of an agent-based model simulation

2.2. Image processing methods

Experimental observations and evidences are required before the development of an agent-based model. It is important to first understand the

experimental observations and then properly design an algorithm accordingly. Thus quantitative approaches are needed to analyze the experimental data and obtain useful information for the development of the agent-based model. For agent-based modeling of *M. xanthus*, the useful information includes possible social interactions between cells, motility of individual cells and the self-organization patterns. Quantitative methods are also required to do proper comparison between experimental results and simulation results for verification of the agent-based model. In studying the self-organizing patterns of *M. xanthus,* microscopic images are an important experimental data source. There are two types of microscopic images. One is grayscale images representing cell densities that could either come from a brightfield microscope or a DIC microscope. The grayscale images provides pattern information of *M. xanthus* swarms. The other is fluorescent images highlighting individual cells that come from fluorescent microscope. The fluorescent images provide data about individual cell behaviors and intercellular interactions. This section contains the discussion about various methods that are used for analyzing these images.

2.2.1. Time-lapse microcinematography and image processing

Time-lapse microcinematography was popular in recording the dynamics of a bacteria colony on a plate for a period of time [37, 118]. The output of the time-lapse microcinematography is a set of images with equal time interval. In the two studies of fruiting body formation dynamics and the fate of developing aggregate of

M. xanthus, different image processing techniques are adopted in adaption to the images from different sources.

2.2.1.1. Image segmentation using intensity threshold and size threshold

Images in this analysis come from time-lapse microcinematography on developing swarms of *M. xanthus* DK1622 as previously described[37, 118] with the following modifications: images were acquired using brightfield microscopy with 20X magnification; images were acquired every minute for 24 hours; images were saved as sequential files and processed as a group. Aggregates and fruiting bodies appear as dark blobs within the developing swarm.

Sequential time-lapse microcinematography images were processed in several steps to enable the automated identification and measurement of aggregates using image-processing capabilities of Matlab (MathWorks). First, each image was divided into 100 quadrants (10 by 10) so that any inconsistency in light intensity across the image could be isolated and its effect on aggregate identification minimized. The contrast of each quadrant was then adjusted separately before the image was reassembled. A gray scale threshold filter was then applied to each quadrant, and a value corresponding to approximately 60% of the mean intensity of the sub-image was selected to compensate for differences in exposure in one image and between images within the same stack. In the resulting binary image, pixels with value of '1' indicate regions of high cell density, and therefore represented potential aggregates. To remove noise, a size threshold of 300 pixels (an area of 0.004mm2, or about 1/5 of average fruiting body, or equivalent diameter of

approximately 72 μm) was applied, and smaller areas were discarded. The remaining pixel clusters with value '1' were automatically assigned consecutive numbers, so that the total number of clusters corresponded to the total number of aggregates in each image. In some image stacks, small dark regions were present from the first frame, long before the onset of development; these were likely due either to small clumps of cells that were not adequately dispersed in liquid culture or to particles (i.e. dust, hair, etc.) that were trapped during the assembly of the microcinematography apparatus. To prevent this source of experimental error from interfering with the automated analysis of aggregates, these regions were eliminated during processing. Some of these dark regions resolve during the late development so they were eliminated only for the early images.

The position of each aggregate was reported as its center, defined as the center of mass of its binary image (i.e. each set of coordinates x and y are respectively equal to the mean of x's and y's for all the pixels within the cluster). For each image, 100 randomized images were generated *in silico* as a means of measuring order in the experimental distribution of aggregates, and a comparison of aggregate distributions between random and experimental images was used to determine if the observed distribution was ordered. To produce randomized images from each experimental image, each aggregate was shifted by a random offset and rotated by a random angle. The new position and orientation of each aggregate was not allowed to overlap with the other shifted and rotated aggregates; if an overlap occurred, then a new shift and rotation was generated, and this process was repeated until there was no overlap.

2.2.1.2. Image segmentation algorithm based on Markov random field

Image segmentation aims to separate aggregates from one another and from the inter-aggregate spacing "background" based on their gray-scale intensities. The processing starts with illumination correction to make the background intensity even throughout the image. Following a crude segmentation, a smooth surface is fitted to the background [119] and then subtracted from the original data to get the illumination-corrected image. Subsequent segmentation was based on consecutive application of a Markov-random field (MRF) segmentation algorithm and morphological operations. MRF segmentation was solved using an approximate algorithm based on belief propagation as in [120]. The initial segmentation by MRF contains multiple aggregates that are very small as well as aggregates that are connected by narrow segments. To solve this problem IPT (Image Processing Toolbox of matlab) functions *imopen* is applied with parameters to remove connections less than 11 pixels across and *bwareaopen* with parameters to ignore aggregates and background regions with fewer than a threshold number of pixels (50 and 300 pixels respectively). As a result some false detections and false connections are removed. Since these operations may also remove the true targets the resulting segmentation is used as an initial condition and performed a second iteration MRF segmentation. In the final adjustment step, *bwareaopen* functions (with a threshold of 300 pixels for both aggregate and background patches) were employed to remove small objects. Such two-step processes work reliably to detect aggregates at the late stages of aggregation. To ensure that segmentation errors do not significantly affect the analysis a *starting frame* was chosen so that the variation

in the numbers of aggregates was limited frame-to-frame (the running variance with 15 frames no more than 5 aggregates). The chosen time of the starting frame corresponds to ~13.5 h of development and depicts a time at which quasi-stable aggregates reliably detectable by the proposed algorithm are formed.

During image segmentation, the movie is processed frame by frame. Each frame image is segmented into foreground (aggregates) and background. The segmentation process has five steps: preprocessing (lighting correction), first Markov random field (MRF)-based segmentation, morphological operation (open), second MRF segmentation, and final adjustment. A workflow illustration is shown in Figure 2-6, and each step is explained in detail as following.

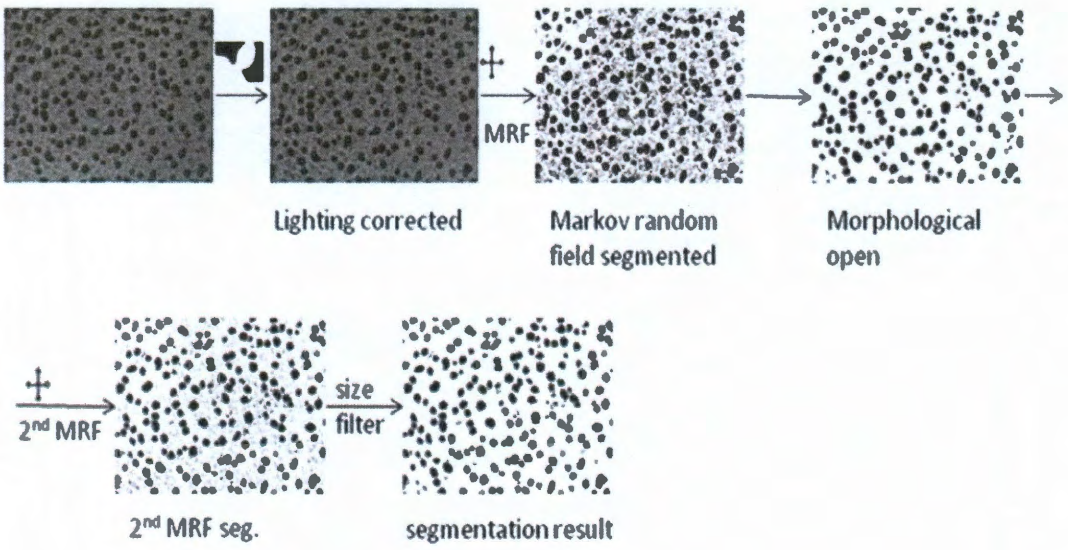


Figure 2-6 The workflow of the two-stage MRF-based image segmentation process

Lighting illumination correction. The segmentation stage aims to separate aggregates from one another and from the inter-aggregate background based on

grayscale intensities. The processing starts with illumination correction to make the background intensity even throughout the image. It is assumed that the illumination distortion generates a smooth surface. Therefore, the idea of illumination correction is to fit a smooth surface on the background pixels and then, subtract this illumination surface from the original image to get the illumination-corrected data. The entire image is 1,200 × 1,600 pixels. First, divide the whole image into small patches of 200 × 200 pixels. On each patch, use the IPT function *graythresh* to find a threshold and then, increase the threshold by 20% to reduce the contribution of segmentation errors to the background. Next, the resulting threshold intensity is used to separate the image patch into foreground and background. A background mask is then obtained from the combination of all of the patches and used to extract background pixels from the whole image. A smooth 2-D surface is then fit to the resulting background intensity with the third-party Matlab function package *gridfit*[121]. This surface is, thereafter, subtracted from the original image.

First MRF segmentation. The MRF-based segmentation method is applied to segment the aggregates from the surrounding area. Markov random field models provide a robust and unified framework for segmentation problems. The segmentation is a labeling process in which a label $f_p \in L$ is assigned to each pixel $p \in P$. It is assumed that the labels should vary smoothly almost everywhere but may change dramatically at some places such as object boundaries. The quality of labeling is given by an energy function that will be minimized. Finding a labeling with minimum energy corresponds to the maximum a posteriori probability

optimization problem. To approximately solve this optimization problem, the fast belief propagation approach [120] is applied to define the MRF energy function. As a result, an approximate labeling with a minimum cost of the energy function is obtained. The segmentation algorithm is implemented in Matlab. To accelerate the process, the illumination-corrected image is divided again into small patches of 200 × 200 pixels, applied *graythresh* to find a threshold, and this threshold is used to get an initial binary segmentation of aggregates and background. This initial segmentation is used as the initial condition for MRF and applied to each patch.

Morphological opening operation. The initial segmentation by MRF contains multiple aggregates that are very small as well as aggregates that are connected by narrow segments. To improve uniformity, *imcomplement* is used to make the aggregates white for the foreground, and then, IMT function *imopen(bw,ones(11))* is used to remove connections less than 11 pixels across, *bwareaopen(bw,300)* is used to ignore aggregate regions with fewer than 300 pixels, and *imcomplement* is used to reverse the image. *bwareaopen(bw,50)* is applied to ignore background regions with fewer than 50 pixels. As a result, false detection and false connections are removed.

Second MRF segmentation and final adjustment. Because these operations may also remove the true targets, the resulting segmentation is used as an initial condition and a second iteration MRF segmentation is performed. The final adjustment is also a morphological operation step, with *bwareaopen(bw,300)* used on both background and foreground to remove small objects. Such two-step

processing reliably detects aggregates at the late stages of aggregation. To ensure that segmentation errors do not significantly affect this analysis, a starting frame was chosen so that the variation in the numbers of aggregates was limited frame to frame (the running variance with 15 frames was no more than five aggregates). The chosen time of the starting frame corresponds to ~13.5 h development and depicts a time at which quasistable aggregates reliably detectable by the proposed algorithm are formed.

The segmented images are, thereafter, converted to binary images, with black aggregates surrounded by the white spacing. Segmentation is applied to each time frame image and then linked with subsequent images by detecting overlapping aggregates. This approach allows us to discern the three aggregate fates depicted.

2.2.2. Feature extraction and quantification

Statistical analysis of the images is used to decrypt properties of individual aggregates that influence their fate. To this end, an extensive list of features (parameters) is devised that can characterize each aggregate according to its geometry, proximity to other aggregates, and size. A complete list of the 33 features is automatically identified for each aggregate present in each frame of the time-lapse movie (Table 4). The list of the features introduced is quite exhaustive and sufficient for the purposes but can, nevertheless, be further extended with new features or mathematical functions of existing features. The features identified in the list are not completely independent and may show a high degree of correlation. To further identify biologically meaningful features of aggregates, the list of 33

features for <160 aggregates from the last time frame of the aggregation movie are automatically identified and used to compose 33 vectors (one for each feature) containing the values of these feature for each of the aggregates. The Spearman correlation between the features was calculated, and Euclidean distance between correlation coefficients was used to build a binary tree-based clustering of the features. The features divide into four major clusters representing features associated with the aggregate's proximity to neighbors (1–3), various size parameters (4–15), parameters of the aggregate's neighbors (17–28), and image shape and topology (30–33). It is interesting to note that aggregate orientation (29) is clustered together with shape-related features. However, feature 16, the ratio of the nearest neighbor distance to the equivalent diameter, clusters separately from both proximity and size. To extract quantify features, it is necessary to link the same aggregates from frame to frame.

2.2.2.1. Link aggregate frame by frame

The aggregate linkage stage aimed to track aggregate movement, shrinkage and expansion, frame-by-frame. Beginning with the aggregates detected at the starting frame, they are numbered, and traced down until disappear or the end of the movie. Aggregates on two continuous frames are then linked based on displacement of the aggregate center being smaller than 23 μm (10 pixels) and overlap in aggregate pixels being larger than 10%. The process is repeated until the last frame of the movie. This process automatically detects newly emerging aggregates and the aggregates that disperse, merge or split.

The linkage algorithm is developed based on two observations. First, the initial aggregates can merge, split, or disperse, and new aggregates might appear. Second, aggregates do not move or only move a little bit from one frame to the next (5 min real time). Each aggregate in the starting frame has been numbered, and the corresponding aggregates on the following frames are traced and recorded until they disperse or until the end of the movie. The linkage method is based on the overlap and the distance between the centroids of aggregates in different frames. The logic is briefly summarized as follows:

i. If an aggregate has less than 10% overlap with any aggregate in the previous frame, a newly appeared aggregate is assumed and added to the list.

ii. If an aggregate overlaps with some aggregates in the previous frame, then one of the following three scenarios occurs.

   a) If the distance between the centroids of this aggregate and the overlapped aggregates in the previous frame is less than a threshold of 23 μm or 10 pixels, then this aggregate is assumed to be the same aggregate as the one in the previous frame, and it is recorded at a new time index as the same one with an updated centroid location and area size.

   b) If the distance is larger than the threshold, splitting or merging is assumed to occur. In this case, a new merged or two new split aggregates are added to the list at this time index with their centroids,

area size, and a pointer directing to the original aggregates from the initial list.

c) The original aggregates at this time index are also updated to indicate merging if the number of overlaps is one or splitting if the number of overlaps is more than one.

Finally, based on the linkage map, all aggregates on the starting frame are labeled as dispersing, merging, or splitting. A segmentation result with a color-coded label is shown in Figure 2-7. Red indicates steady aggregates, yellow indicates dispersing aggregates, blue indicates merging aggregates, and cyan indicates splitting aggregates. The linkage is manually reviewed and curated to ensure that the algorithms function properly, especially for relatively rare events of splitting and merging.

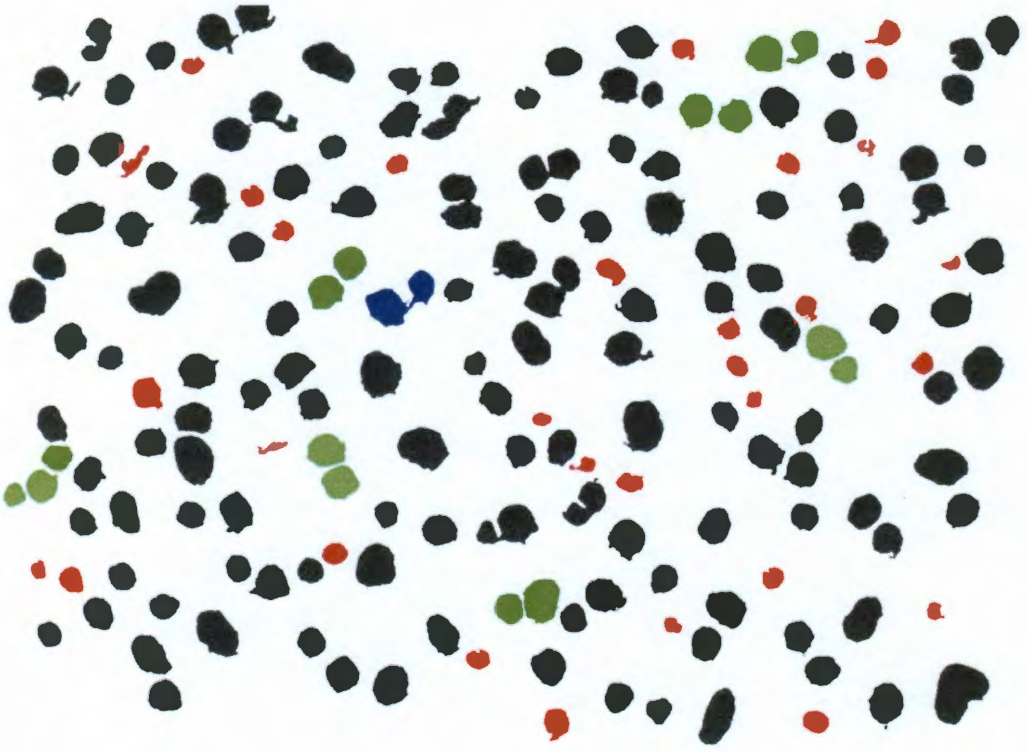


Figure 2-7 A segmentation result of an aggregation image at 13.5 h with color-coded coloring (black, stable aggregates; red, dispersing aggregates; green, merging aggregates; blue, splitting aggregates)

2.2.2.2. Feature quantification

After segmentation and linkage, aggregates on a specific frame are labeled as dispersing, merging, splitting, or stable. Feature extraction is performed on the 2D black and white image. Matlab IPT function *regionprops* measures a set of properties for each connected component in the binary image. Each segmented aggregate is a connected component, and *regionprops* is used to extract image features for aggregate fate analysis. The following is a list of eight features (the feature index number is the same as that inTable 4) from the *regionprops* properties.

1) Perimeter: scalar; the distance around the boundary of the region (i.e., a segmented aggregate). 2) Equivdiameter: scalar that specifies the diameter of a circle with the same area as the region. 3) Area: scalar; the actual number of pixels in the region. 4) Orientation: scalar; the angle between the x axis and the major axis of the ellipse that has the same second moments as the region. 5) Equivalent diameter/perimeter. 6) Solidity: scalar specifying the proportion of the pixels in the convex hull that are also in the region. 7) Eccentricity: scalar that specifies the eccentricity of the ellipse that has the same second moments as the region. The value is between zero and one. 8) Minor axis length/major axis length: minor or major axis length is a scalar specifying the length (in pixels) of the minor or major axis of the ellipse that has the same normalized second central moments as the region.

Regionprops property centroid provides a vector that specifies the center of mass of the region (i.e., an aggregate). The displacement relationship of all aggregates in the field can be extracted based on their centroids. Then, the nearest neighbor (NN) of a target aggregate and its immediate neighbors at various directions around the target aggregate can both be extracted. Figure 2-8 illustrates the relationship of the target aggregate (TA), the nearest neighbor (NNA), and the immediate neighbors at various directions around the target (1, 2 . . . 5).

The following is a list of 25 features (the feature index number is the same as that in Table 1) calculated based on the NN or immediate neighbors.

1) Distance to the NN/average weighted area of neighbors (i. e., weighted sum of all areas of neighbors within a circular neighborhood with one/distance being a weight). 2) Average distance to neighbors. 3) Distance to the NN. 4) Equivalent diameter/distance to the NN. 5) Area/area of the NN. 6) Area/neighboring area. 7) Area/sum of the area of neighbors. 8) Equivalent diameter/average weighted area of neighbors. 9) Area/maximal area of neighbors. 10) Area/median area of neighbors. 11) Area/average area of neighbors. 12) Area/minimal area of neighbors. 13) Distance to the NN/equivalent diameter. 14) Neighboring area (area of circular neighborhood filled with aggregates). 15) Sum of distance to neighbors. 16) Weighted sum area of neighbors. 17) Sum of equivalent diameter of neighbors. 18) Sum of area of neighbors. 19) Area of the NN. 20) Minimal area of neighbors. 21) Average weighted area of neighbors. 22) Median area of neighbors. 23) Maximal area of neighbors. 24) Average equivalent diameter of neighbors. 25) Average area of neighbors.

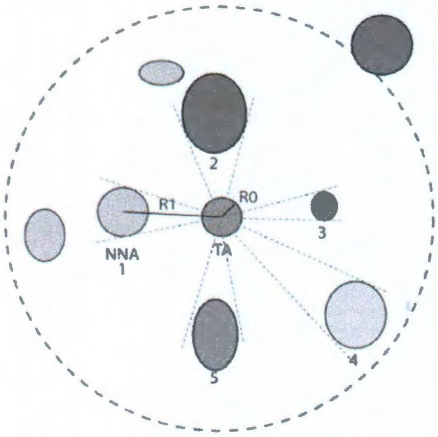


Figure 2-8 The target aggregate (TA), the nearest neighbor (NNA), and the immediate neighbors (1–5) at various directions around the target. Note that, in each direction, only the immediate neighbor is considered.

Data analysis begins with automatic extraction of the features of each aggregate Table 4. The IPT function *regionprops* computes several aggregate properties. Algebraic manipulations of its outputs lead to all 33 features of each

aggregate for each frame. Having features and fate labels for all aggregates, information theory and machine learning approaches are is employed to connect features and fate [122]. Normalized mutual information, $NMI = I(D,F_i)/H(D)$ was used to find single features that are highly interdependent with aggregate fate Eq. (8) and Eq. (9). In Eq. (8) D is a boolean random variable with a value d=1 for dispersed and d=0 for stable aggregate and random variable F_i is for each of the 33 features identified. $p(d)$ is the probability distribution of aggregate dispersal, $p(f^i)$ is a probability distribution of the given feature (i=1,...33) whereas $p(d,f^i)$ is a joint probability distribution between the dispersal and a given feature. In Eq. (9), *H(D)* represents uncertainty (entropy) of aggregate fate whereas $H(D|F_i)$ is a conditional uncertainty of the fate given the feature value. If a given feature is a good predictor of the aggregate fate, conditional uncertainty is small and $I(D,F_i) \approx H(D)$. On the other hand, if feature values are independent of fate then $p(d,f^i) \approx p(d)p(f^i)$ and $I(D,F_i) \approx 0$. Support Vector Machine analysis (based on MATLAB *svmtrain* and *svmclassify* commands) were employed to test whether multiple feature combinations can better predict the fate of an aggregate [123].

$$I(D,F_i) = \sum_{d\in D}\sum_{f^i\in F_i} p(d,f^i)\log\left(\frac{p(d,f^i)}{p(d)p(f^i)}\right) \qquad \text{Eq. (8)}$$

$$I(D,F_i) = H(D) - H(D|F_i);\;\; H(D) = \sum_{d\in D} p(d)\log\left(\frac{1}{p(d)}\right);$$
$$H(D|F_i) = \sum_{f^i\in F_i} p(f^i)\sum_{d\in D} p(d|f^i)\log\left(\frac{1}{p(d|f^i)}\right) \qquad \text{Eq. (9)}$$

Table 4 Aggregate features

#	Description
1	Distance to the NN* / average weighted area of neighbors#
2	Average distance to neighbors
3	Distance to the NN
4	Equivalent diameter / distance to the NN
5	Area / area of the NN
6	Area / neighboring area
7	Area / sum of the area of neighbors
8	Equivalent diameter / average weighted area of neighbors#
9	Area / maximal area of neighbors
10	Area / median area of neighbors
11	Area / average area of neighbors
12	Area / minimal area of neighbors
13	Perimeter
14	Equivalent diameter, $2 \cdot \sqrt{Area / \pi}$
15	Area
16	Distance to the NN / equivalent diameter
17	Neighboring area+
18	Sum of distance to neighbors
19	Weighted sum area of neighbors
20	Sum of equivalent diameter of neighbors
21	Sum of area of neighbors
22	Area of the NN
23	Minimal area of neighbors
24	Average weighted area of neighbors
25	Median area of neighbors
26	Maximal area of neighbors
27	Average equivalent diameter of neighbors
28	Average area of neighbors
29	Orientation
30	Equivalent diameter / perimeter
31	Solidity
32	Eccentricity
33	Minor axis length/ major axis length

*Nearest neighbor; #Weighted sum of all areas of neighbors within a circular neighborhood with 1/distance being a weight
+ Area of circular neighborhood -filled with aggregates

2.3. Statistical analysis methods

2.3.1. Cumulative radial distribution function

A radial distribution function is commonly used in statistical mechanics to characterize isotropic distributions in two and three dimensions. It is typically defined as a normalized density of particles in an infinitesimal ring from radius *r* to *r+dr*, but this definition is very sensitive to noise if the number of particles is small, as is the case with aggregates in a developing swarm. To characterize a two-dimensional distribution of aggregates on a swarm, cumulative radial distribution function *(CRDF)*, $g_c(r)$, is used as the measurement. Starting at the center of each aggregate, the circle of radius r was identified and the number of aggregates were counted in the circle, *N(r)*. The aggregate density in this circle is computed as:

$$f(r) = \frac{N(r)}{\pi r^2} \qquad \text{Eq. (10)}$$

Periodic boundary conditions are assumed, i.e. distribution outside the image is the same as the distribution in the image. To normalize, the radial density is divided by the average fruiting body density to compute the *CRDF* $g_c(r)$:

$$g_c(r) = \frac{f(r)}{f(\infty)} \qquad \text{Eq. (11)}$$

For each image, *CRDF* is computed by averaging the radial distribution functions centered at each aggregate. As defined, $g_c(r)$ depends on the size of the aggregates in the image. To partially compensate for this dependence the

distribution is presented as a function of dimensionless radius r/R_0, where R_0 is the radius of aggregates in the image, $R_0 = \sqrt{\bar{A}/\pi}$, and $\bar{A}$ represents the area of aggregates. This scaling eliminates dependence of $g_c(r/R_0)$ on the size of aggregates for the randomized control. For each set of images corresponding to a particular developmental stage, the mean and standard deviation of $g_c(r/R_0)$ for 20 different images is computed.

2.3.2. Aggregate feature clustering

A set of 33 features encompassing multiple aspects of each aggregate was automatically detected for more than 150 aggregates from the last frame of a time-lapse movie. The Matlab statistics toolbox (ST) function *corr* (Spearman type) is used to compute a correlation coefficient matrix between feature values. The Spearman rank correlation assesses how well the relationship between two variables can be described using a monotonic function, linear or nonlinear, to capture the interdependence of two different features such as area and equivalent diameter. The ST function linkage creates an agglomerative hierarchical cluster tree from the correlation coefficient matrix. *linkage(CCoeff,'average','euclidean')* is used to get four major classes and apply the ST function *dendrogram* to draw the clustering tree shown in Figure 4-2. The features divide into four major clusters representing features associated with the proximity of the aggregate to neighbors (1–3), various size parameters (4–15), parameters of the aggregate's neighbors (17–28), and image shape and topology (30–33).

2.3.3. Mutual Information and Support Vector Machine Analysis

Having feature and fate labels for all aggregates, information theory and machine learning approaches are used to connect features and fate (3). Normalized mutual information, $NMI = I(D, F_i) / H(D)$, was used to find single features that correlate with aggregate fate. To estimate mutual information, the probability distribution of aggregate dispersal pðdÞ, a probability distribution of the given feature (i = 1 . . . 33) $p(d)$, and the joint probability distribution between the dispersal and a given feature $p(f^i)$ are all estimated based on 1-D and 2-D histograms using Matlab function *histc*. Support Vector Machine (SVM) analysis was used to test whether multiple feature combinations can better predict the fate of an aggregate [124]. SVMs have been widely suggested for binary classification. The linear SVM defines a hyper plane in the feature space, which separates the training examples of the two classes. The problem of determining the hyper plane can be formulated as a convex quadratic programming problem. If the classes are not linearly separable, by relaxing the constraints and introducing a slack parameter, a similar quadratic programming problem can be formulated and solved. The SVM analysis is based on Matlab bioinformatics toolbox (BT) functions *svmtrain* and *svmclassify*. In the analyzed movie, there are 160 examples of steady aggregates and 91 examples of dispersing aggregates available for training and testing. Each example is represented by a multidimensional (up to 33D) feature vector. In each case, 75% of the data is used to train SVM, whereas the remaining 25% is used to estimate the prediction error rate. The data index is randomly permutated, and the

experiments have been repeated 30 times using different combinations of training and testing data to reduce the training bias and estimate SE. The error rate is defined as the percentage of cases (both false positive and false negative) where dispersal is not correctly predicted by a threshold model. Any single feature or feature combination vector can be used to run the SVM to test whether they predict the fate of transitional aggregates. Because all combinations of the 33 features are too many to test, only a selected set of combinations are tested in this analysis. First, every single feature is tested and the features are sorted based on their average prediction error from the lowest to the highest. The size related feature generates the lowest error rate, consistent with the mutual information analysis. Then, the combination of the top one to k (k = 1 . . . n) features based on the sorting are tested, and the results are shown in Figure 4-3 C.

2.3.4. Analysis of cell motility in florescent microscopic movies

In order to obtain data on individual cell motility information, tracking individual cells for a certain period of time is needed. To this end, 0.5% of all *M. xanthus* cells are labeled with green fluorescent protein (GFP). Time-lapse images (in 1 or 2 min intervals) were taken under florescent microscope for 2-4 hours. Thereafter, ImageJ software[125] is used to track individual cells in all fluorescent movies.

The tracking result of one cell includes coordinates of the cell in a given frame number n: *x(n)* and *y(n).* This information is used to compute motility parameters of cells with following procedures:

Instant cell displacements and speeds: First, the instant displacement of each cell between two consecutive frames is computed as:

$d(n) = \sqrt{(x(n) - x(n-1)^2 + (y(n) - y(n-1))^2}$. Then the instant cell speed is computed as $v(n) = \frac{d(n)}{\delta t}$, where δt is the time interval between successive frames.

Removal of temporary immotile cells: From analysis of the cell trajectory several instances are observed when fluorescent cells cease their motility for extended periods of time (over 5 min). Therefore this data is detected and removed from the trajectories. I look for the continuous time-intervals 5min or longer in which the instant cell speed is less than 0.2μm/min (less than 5% of average speed) and exclude these intervals from the motility statistics.

Detection of cell reversals: In order to calculate statistics of cell reversal periods an algorithm is needed that robustly detects cellular reversals but ignores changes in cell velocity associated with random motion and/or turning. For the experimental setup, the problem is simplified by the observation that cells are predominantly aligned and therefore travel in the same direction. This direction generally speaking is not parallel to image boundaries. Therefore the images are rotated to ensure that directed motility of the most of the cells is along x-axis. To determine the direction which cells are predominantly aligned to, principal component analysis (PCA) is applied to cell trajectories obtained from individual cell tracking. PCA is used in multivariate analysis to determine the important information in data. The variance of data along the first principal component is

always the largest. The first principal component can be rotate to parallel to X-axis. As a result, the most of cell movement is now parallel to X-axis.

(1) Obtaining the average direction of cell movement.

When all cells' centered coordinates are placed together, it is clear that their trajectories are mostly along one direction (Figure 2-9). These directions are not necessary align with the microscope images boundaries. Therefore, the first step is to rotate cell coordinates so that the directed motility of cells occurs along the new x-axis.

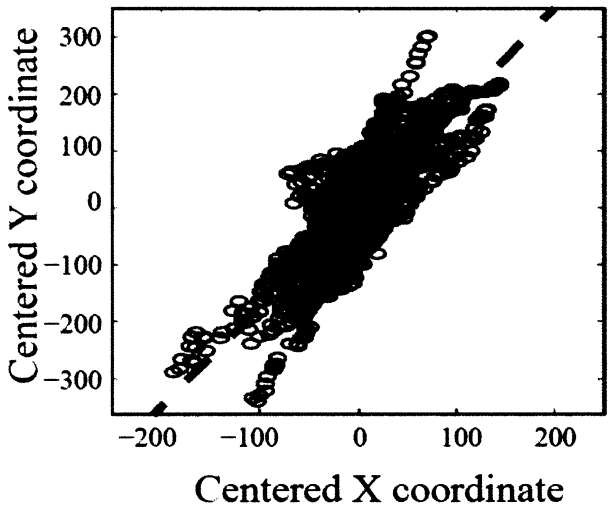


Figure 2-9 All the cell coordinates are centered by subtracting the average position of each cell. Then trajectories of all cells are put together and apply PCA. The dash line is the regression line.

x_i and y_i are used to denote the coordinates of points in cell i's trajectory in the original images. First, each cell trajectory is centered by subtracting its mean:

$x_i' = x_i - \langle x_i \rangle, y_i' = y_i - \langle y_i \rangle$. Then all cell coordinates are put together to form the row vector of x and y: $x' = (x_1', x_2', ..., x_N'), y' = (y_1', y_2', ..., y_N')$.

Principal component analysis (PCA) is applied to obtain the direction that cell coordinates have the largest variance that is termed as the first principal component. This direction will be the new x-axis. The centered coordinates are combined into a matrix *XY* as

$$XY = \begin{pmatrix} x_1' & y_1' \\ x_2' & y_2' \\ & \\ x_N' & y_N' \end{pmatrix}.$$

To compute the principal components, the covariance matrix is needed of the coordinate data:

$$\mathrm{cov}(XY) = \begin{pmatrix} \mathrm{cov}(x', x') & \mathrm{cov}(x', y') \\ \mathrm{cov}(y', x') & \mathrm{cov}(y', y') \end{pmatrix},$$

where the covariance is defined as:

$$\mathrm{cov}(x', y') = \frac{\sum_{i=1}^{N} (x_i' - \overline{x'})(y_i' - \overline{y'})}{N-1}.$$

Next the eigenvalues and unit eigenvectors corresponding to the eigenvalues of the covariant matrix *cov(XY)* are computed and used to construct a rotation matrix *EV*, a 2 by 2 matrix with a unit eigenvector in each column: $EV = \begin{pmatrix} eig_1 & eig_2 \end{pmatrix}$, where eig_1 represent a normalized eigenvector with the largest eigenvalue, eig_2 is

the second normalized eigenvector. This matrix can be used to rotate cellular coordinates: $XY^T = EV^T \; XY^T$, where

$$XY = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ & \\ x_N & y_N \end{pmatrix}$$

contains rotated coordinates with majority of directed motility occurring along its first component.

(2) Capturing the sign changing points in the new coordinates.

Most of cell movements are along one direction which was computed above using PCA. As a result, the search for reversal points in cell trajectories can be based on the sign changing points in the time-derivative of the coordinates along the first principal component(x_i). The sign of a coordinate *s* is defined as: value of +1 means increasing in coordinate values and value of -1 means decreasing in coordinate values $s(n) = \text{sign}(x_i(n) - x_i(n-1))$. When s=+1 given cell travels along the principal axis in the positive direction whereas s =-1 it travels in the opposite direction. Here the $x_i(n)$ denotes the cell *i*'s new coordinate at time *n* and *s(n)* is the sign of the coordinate. Then the sign changing vector *sc* can be calculated as*:*

$sc(n) = s(n) - s(n-1)$.

All the non-zero points in *sc* are potential reversal points since the changing sign means cells switches its direction of motion along the principal axis. However,

not all of the sign changing points are the real reversal points as noise in cellular motility and tracking can result in false-positives. Therefore, additional analysis is applied on the sign changing data *sc* to find the real reversal points.

(3) Obtaining real reversal points in $x(n)$ from the sign changing points

There are several sign changing points that occur consecutively during a very short period of time. The cell behaviors during these periods are referred as "tumbling". The following algorithm describes how these "tumbling" events are dealt with. For illustration, Figure 2-10 shows sample application of the algorithm

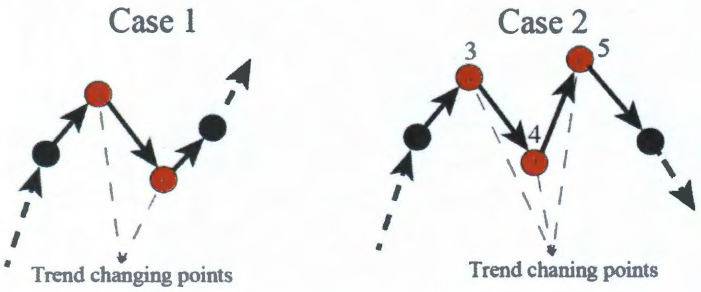


Figure 2-10 A schematic diagram showing two types of consecutive cell reversals. In one type the cell changes direction finally and in the other the cell continues with the same direction.

i. The range of indices with successive non-zero values in *sc* is identified. These are the tumbling events as the example shown in Figure 2-10. The red dots in each case form a tumbling event.

ii. The number of sign changing points are then counted for each tumbling event. For example, in Figure 2-10, there are 2 sign changing points in Case 1 and 3 in Case 2.

iii. The "tumbling" events are separated into 2 cases. In Case 1, cell continues the same direction after the "tumbling" as they enter the "tumbling". In other

words, the number of sign changing points in the "tumbling" event is even. For example, Figure 2-10 left panel. Case 2, cells change their direction after the "tumbling" event from the direction they enter the "tumbling". In other words, the number of sign changing points in the "tumbling" event is odd. The schematic diagram is shown in Figure 2-10 right panel.

iv. For case 1, since cells do not change their direction before and after the "tumbling", therefore, all the sign changing points in the "tumbling" event are removed and no reversals occur during that event. So in Figure 2-10 left panel, there is no actual reversal.

v. For case 2, since cells change their direction after the "tumbling" event, all the sign changing points in the "tumbling" event are treated as one reversal. To assign one of the tumbling even points as a reversal point, a point is identified with maximal X coordinate for reversals in which cell was going in positive direction along x (*s=+1*) axis before the reversal and minimal X coordinate if cell was going in the negative direction before the reversal(s=-1). For example, in Figure 2-10 right panel, point 5 is assigned as the actual reversal point.

The reversal period is defined as the time interval between two adjacent reversal points. Suppose a cell reverses at time *i* and its next reversal occurs at time *j*, then the reversal period τ for the cell between time *i* and *j* can be calculated as: $\tau = (j - i)\delta t$.

Selection of rippling cells: Under experimental conditions, even inside the prey region, not all *M. xanthus* cells have rippling behavior. Thus it is necessary to exclude non-rippling cells to enhance the statistical results. Therefore, a fraction of cells are manually chosen to be rippling cells based on the following criteria. (a) The time is counted in which a cell keeps a low motion state, i.e. the instant speed is smaller than 0.5 μm/min. If a cell has a low motion time period larger than 5 minutes, it is excluded from ripping cells. (b) After the rotation of the cell trajectories by principle analysis (PCA), most of the movement of a cell that participates in rippling is along the first principal component. Therefore, rippling cells are chosen if the movement along the first principal component is significantly larger than the movement along the second principal component:

$$\frac{\left|\max(x_i)-\min(x_i)\right|}{\left|\max(y_i)-\min(y_i)\right|}>3.$$

2.3.5. Analysis of gray scale images that represent cell densities

As rippling is essentially traveling cell density wave, it is important to acquire cell density information from both experiments and ABM simulations. In experiments, the *M. xanthus* colony under fluorescent microscope was recorded. The fluorescent images coming from microscope (lower than 20 X magnifications) are actually the representation of local cell density information where the bright represents high cell density and dark represents low cell density. Therefore, the fluorescent images can be treated as gray scale images that represent local cell densities. In ABM simulations, the simulation domain is covered by 1 μm by 1 μm

grid, and the local cell densities are calculated based on the agents whose centers are in each grid square. In the density images, similar to the experimental gray scale images, the bright represents high cell density and dark represents low cell density. *I* is used to denote the gray scale images from both ABM simulations and experimental observations. And *I(x,y)* denotes the pixel value, which represents the local cell density at position *(x,y)*.

Creation of cell density space-time image: The cell density space-time plot shown in Figure 5-3 E is made out of the gray scale fluorescent images from experimental simulations of cell-density images from the simulations. Suppose a cell density image *I* has size of *X*×Y. As most of the cell movements are on the first principal component, therefore the image *I* is rotated so that the first principal component of cell coordinates in image *I* is parallel to x-axis. To rotate image *I*, a large enough image *I'* is needed so that it can contain all the pixels of image *I* after rotation. For each pixel *I'(x,y)* in image *I'*, its corresponding coordinate in image *I* is computed before rotation:

$$\begin{pmatrix} x \\ y \end{pmatrix} = (EV^T)^{-1} \begin{pmatrix} x \\ y \end{pmatrix}$$

The value of the nearest pixel to (x, y) in image *I* is used as the pixel value of *I'* (x, y). If (x, y) is not inside image *I*, $I'(x, y)$ is set to 0. To obtain the 1-D averaged density I_x, columns of image *I'* are averaged over: $I_x = \langle I'(x, y) \rangle_y$. The brackets denote average over pixel values that were part of the original image. Next I_x is

detrended by subtracting the least-square fit of it from itself. As a result, the 1-D averaged and detrended cell density of I_x is calculated. By putting I_x of different time-steps together a 2-D space-time image of cell density is generated. As the last step, the gray scale of the image is flipped so that the dark color represents high cell density and bright color represents low cell density. As a result, in the space-time cell density image, the dark ridges represent wave crest (high cell density). The time and position of cell reversals are then superimposed on this plot and resulting diagram is shown in Figure 5-3 E. The same procedure is used to produce the space-time plot of simulation results and the result is shown in Figure 5-5.

One hypothesis about rippling is that most of the rippling cells reverse during ripple wave crest collisions. To test this hypothesis by the experimental data analysis, the time and positions of cell reversals in section 2.3.4 and the space-time image of cell density are combined. To test this hypothesis using the ABM simulations, the time and positions of cell reversals are also recorded in the ABM simulations. First, process is done on the space-time images of cell density of both experimental data and ABM simulation data. As the space-time images are always detrended, the pixels with negative values are discarded (by setting to 0). As a result, space-time images with only the dark ridges presented are obtained. The next approach is to identify the collisions of wave crests in the space-time images. From experimental observations of both experimental data and ABM simulation data, each collision area of two wave crests always contains at least one pixel that has the largest cell density (darkest in color) in a local area in the space-time image with wavelength in space and reversal period in time since the overlapping of two

wave crests always leads to higher cell density. As a result, the local maxima in space-time images in a local area of size of wavelength times reversal period is searched for. After the desired pixels are found, a square region is defined centered at each pixel. The region is of size of twice the wave crest width times 3 minutes and represents the collision region of two opposing crests. As the collision region contains two wave crests, the size of this region in space must be at most twice the wave crest width. Given the average wave crest width is ~10 μm, the maximum possible distance a cell travels during two wave crests collision should be ~20 μm, considering that the average cell velocity is 6 μm/min, as a result, the size of the collision region in time is set to be 3 minutes. Next the time and position of each cell reversal is went through and checked whether or not it is in any of the wave collision region. To obtain the variations bootstrapping method is used to calculate the mean and standard deviation of the percentage of collisions inside wave collision area. The analysis reversals that 75.0% (±2.6%) of all tracked reversals in prey area happen during wave crests collisions in experimental observations. In the ABM simulations 82% (±2.1%) of rippling cells reverse during wave crest collision. By adding 10% non-rippling cells in simulations, only 17% (±2.3%) of non-ripping cells reverse during wave crests collision.

Estimation of wavelength and wave crest width: In both experimental observations and ABM model simulations, movement of *M. xanthus* cells are predominantly on the first principal component computed above. As a result, the direction of movement of ripple is also on the first principal component. Thus, the 1-

D continuous wavelet transform (CWT) is used to obtain the wavelength[126-129]. Given a function (or signal) $f(x)$, a mother wavelet $\psi(x)$, the CWT is defined as:

$$c(a,b) = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{a}} \psi^*(\frac{x-b}{a}) dx \qquad \text{Eq. (12)}$$

where a is the scale parameter and $a > 0$; b is the position parameter; $c(a,b)$ denotes the wavelet coefficient computed from Eq. (12) at scale a and position b; * denotes the complex conjugation.

In 2.3.4, the original experimental images are rotated such that the first principal component of the images is parallel to the x-axis. In ABM simulations, agents are set aligned along the x-axis initially and alignment rule keeps the agents align during simulation. As a result, in both images from experimental observations and ABM simulations, the direction of ripples is along the x-axis. Suppose image I has size of $X \times Y$, in which ripples travel along the x-axis. Then the CWT is applied to each row of the image: $I_i(x) = I(:,i)$, where $i = 1, 2, ..., Y$. To apply the CWT to $I_i(x)$, an assumption is needed that $I_i(x)$ is the result of sampling of a continuous function.

There are many mother wavelet functions, among which Morlet wavelet is chosen:

$$\psi(x) = e^{-x^2/2} \cos(5x).$$

There are several reasons to choose Morlet wavelet over other wavelet functions. The Morlet wavelet can be viewed as a sinusoid wave multiplied by a Gaussian window. The Morlet wavelet is often used when analyzing signal with certain period

since the wavelet transform using Morlet wavelet can be treated as the Fourier transform (Cosine transform) of the signal times a Gaussian window as the Fourier transform (Cosine transform) of the Gaussian window is still a Gaussian window [130, 131].

As a result, each $I_i(x)$ has a corresponding $c_i(a,b)$ using Eq. (12), and the next step is to average the wavelet coefficient along all i (along y-direction):

$$c(a,b) = \frac{1}{Y}\sum_{i=1}^{Y} c_i(a,b)\,.$$

As shown in Appendix figure 1 panel A and C, several peaks are presenting in the absolute value of wavelet coefficient. Wavelengths are calculated by the distances along X-axis (horizontally) between every other peak in the absolute value of wavelet coefficients. In order to do that, $c(a,b)$ is averaged over the scale parameter a:

$$c(b) = \frac{1}{N_0}\sum_{a=1}^{N_0} c(a,b) \qquad \text{Eq. (13)}$$

where N_0 is the largest scale used in the wavelet transform. Suppose in $c(b)$ there are N peaks with position parameter $b_1,\ldots,b_N$, the average wavelength λ of one image is calculated by:

$$\lambda = \frac{\sum_{i=1}^{N-2}(b_{i+2} - b_i)}{N-2} \qquad \text{Eq. (14)}$$

Eq. (14) is used for ABM simulation images. For fluorescent images, small changes are made to the equation to omit the peaks at the two ends b_1 and b_N. The reason is that the peaks at b_1 and b_N are deeply affected by the discontinuity at the edges of the fluorescent images. Periodic boundary condition is used for ABM simulation so there is no effect of discontinuity at the edges. The wavelengths from different frames are then averaged to get the averaged wavelength. 20 fluorescent images and 20 ABM simulation images are used to calculate these data points.

To determine the wave–crest width similar procedures are performed but assume the wave-crest is well-approximated by a Gaussian distribution

$$N(x,\sigma) = e^{-\frac{x^2}{2\sigma^2}}.$$

Therefore, the best fitted Gaussian distribution is searched for and used the width σ as the estimation of the wave crest. First, starting with 1-D densities $I_i(x)$, the moving correlation coefficients of Gaussian function and $I_i(x)$ is calculated:

$$f_i(\sigma,b) = \int_0^X I_i(x)\frac{1}{\sigma}N(x-b,\sigma)dx.$$

Then the correlation coefficients are averaged:

$$f(\sigma) = \frac{1}{XY}\sum_{b=1}^{X}\sum_{i=1}^{Y} f_i(\sigma,b),b).$$

And the wave crest width Δ can be obtained by:

$$\Delta = \arg\max_{\sigma} f(\sigma) \qquad \text{Eq. (15)}$$

2.3.6. Estimating colony expansion in experimental images

ImageJ software is used to obtain *M. xanthus* colony edge information from analysis of gray scale images from DIC microscope in experimental observations. For each image, points are manually picked on colony edge and their positions are recorded. The number of points should be enough that the linear interpolation of these points can sufficiently represent the real colony edge. The linear interpolation is used to connect these points and as a result a curve is obtained that represents the colony edge. To calculate the movement of edge from image to image, the following procedure is applied:

(1) Suppose there are two curves of edges of the same colony from two different images, curve 1 and curve 2. For each line segment in curve 1, the position of the mid-point of the line segment is computed by averaging the coordinates of the two end points.

(2) The distance from the mid-point of each of the line segment in curve 1 to curve 2 is then computed and recorded.

(3) The movement of colony from one image to the other is the average distance of the distances calculated in step (2).

Chapter 3

Quantifying aggregation dynamics during *Myxococcus xanthus* development

3.1. Introduction

When a vegetative growing *M. xanthus* colony deprive all the available nutrients, cells in the colony stop their expansion and redirect their movement inward into multiple location where they form aggregates of cells later on. After several hours, many dome-shaped aggregates of *M. xanthus* cells are formed with size around 0.1 mm in diameter and could contain up to 1 million cells. Under standard experimental conditions, a typical swarm of several million cells will produce hundreds of fruiting bodies over an area of less than 1 cm^2. As an important self-organizing pattern during development, aggregation process of *M. xanthus* is intensively studied by both biologists and theoretical modelers [13, 14, 21-23, 28,

32, 33, 39, 41, 47, 90]. Along with the aggregation process, several transient multicellular patterns occur within a swarm during development. Traveling waves [6-8, 78], swirling vortices [25], and multilevel terraces [37] have all been observed prior to and during the formation of aggregates and fruiting bodies. These patterns occur because development is an emergent process, where global order is produced from the numerous interactions of lower-level components, so that intermediate patterns arise through nonlinear interactions as the system transitions to a new stable, ordered state [29]. In developing bacteria such as *M. xanthus*, that stable ordered state has long been recognized to be the fruiting body [84], but a developing swarm typically creates a distribution of fruiting bodies.

Based on the knowledge that there is not yet any quorum sensing molecules discovered in *M. xanthus* swarms, what is the mechanism of behind the aggregation process? Are these aggregates (fruiting bodies) randomly placed on the experimental plates or other surfaces? If the placements of these fruiting bodies are not random, then what is the order and how this order is achieved? As an important complimentary approach in understanding the aggregation pattern,

These questions are the objectives of my research and I adopted several mathematical modeling techniques in my approach to these questions. To quantify the ordering of fruiting bodies placements, I adapted the radial distribution function from statistical mechanics. To reveal the mechanism behind the self-organizing aggregation of *M. xanthus* cells into fruiting bodies, I combine the agent-based modeling of *M. xanthus* swarm and the image processing techniques. Based on the

image processing techniques, some hypothesis of mechanisms of aggregation into fruiting bodies of *M. xanthus* are proposed, these hypothesized mechanisms are implemented in the agent-based model and simulations are run using these model to test if the hypothesized mechanisms are reasonable. Then based on the agent-based model simulation results, a measurement of how close the model simulation is comparing with experimental images is developed. The next step is the examination of the difference between theoretical results and experimental observations, and to link these differences to the hypothesized mechanisms. Further, these hypothesize mechanisms can be modified or just toned with some parameters to make the simulated results closer to experimental results. This cycle goes on until quantitative match between simulation and experiments is reached.

This thesis demonstrates that *M. xanthus* developmental self-organization also occurs at the scale of the entire population of fruiting bodies rather than just a single fruiting body. The results show that the arrangement of fruiting bodies within the swarm is not random and becomes more ordered with time. Furthermore, the process through which this ordered placement occurs is not reproduced in the traffic jam model for the formation of fruiting bodies.

3.2. Results

3.2.1. Analysis of microcinematography data reveals various stages in the aggregation process

To record the process of *M. xanthus* development, images of a swarm spotted on starvation agar were acquired at one minute intervals over a period of 24 hours, resulting in a 'stack' of 1,441 sequential images. Acquisition is stopped at 24 hours because, after this time period and under these conditions, a population of aggregates has largely stabilized. A subset of images from one image stack at different times during development is displayed in Figure 3-1 (top row), together with the corresponding processed binary images used for automated analysis Figure 3-1 (bottom row). The dark spots in both sets of images represent regions of sufficient cell density that are considered aggregates.

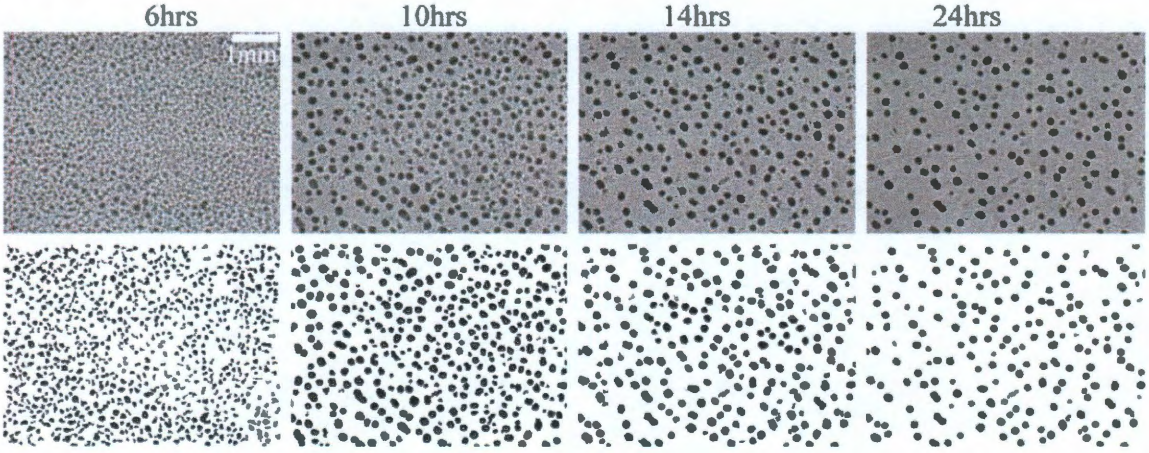


Figure 3-1 Progression of developmental aggregation. Top row: unprocessed images of fruiting body development (6, 10, 14 and 24 hours). Bottom row: corresponding processed binary images; black areas represent aggregates.

The automated quantification of aggregate size, shape, and position is performed on all processed images from each image stack. A plot of aggregate number as a function of time reveals that all aggregates do not follow identical developmental process Figure 3-2 A and B. Aggregates first appear within a few hours following the onset of starvation, and their number rapidly increases until approximately seven hours into development. At this point there is sometimes a brief plateau, followed by a decrease in aggregate number that is almost as rapid as the preceding increase. The rate of decrease slows between 10 and 20 hours, at which point the number of aggregates nearly stabilizes at approximately 25-30% of its peak. This loss of aggregates is evident in both the processed and unprocessed images from Figure 3-1; there are fewer aggregates in each image past the six-hour time point. These data demonstrate that all of the aggregates within a swarm do not undergo the same developmental process. Aggregates have different fates, some grow and mature into fruiting bodies while others grow for a period and then regress and disappear. Aggregate regression and disappearance is clearly visible in the time-lapse microcinematography image stacks.

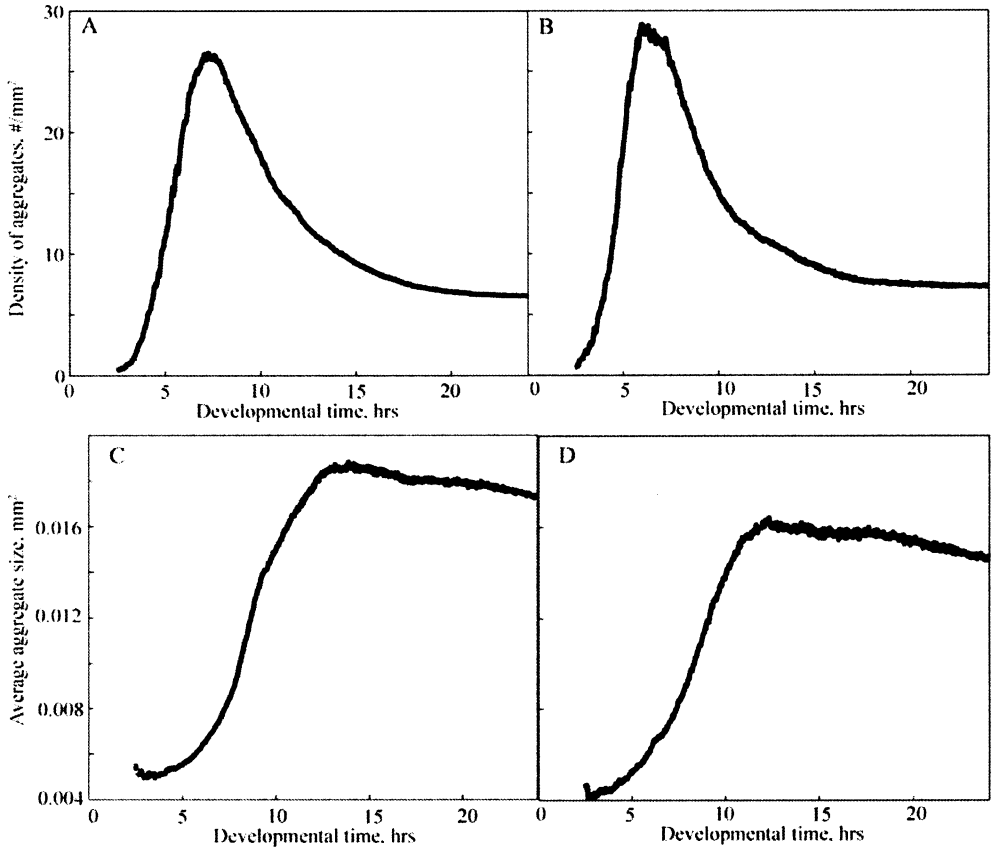


Figure 3-2 Quantification of aggregation dynamics. (A, B) Change in aggregate number over 1mm2 area during development for a single representative experiment (A) and the average of seven replicate experiments (B). (C, D). Change in average aggregate size during a single representative experiment (C) and the average of seven replicate experiments (D).

A plot of average aggregate size as a function of time shows that the 'average' aggregate continues to grow unabated past 10 hours after starvation onset Figure 3-2 C and D. This is significant because, at 10 hours the number of aggregates is rapidly decreasing, but average aggregate size continues to increase. The peak in average aggregate size occurs between 10 and 15 hours, and this is followed by a very gradual decline; this decline may be caused by the slight contraction and sharpening of each aggregate as it matures into a fruiting body. One reason why

average aggregate size continues to increase as aggregate number decreases involves the method of quantification; regressing aggregates reduce the overall average aggregate size, but once an aggregate disappears it is no longer counted, and so the effect of aggregate disappearance opposes the effect of aggregate regression. Another reason is that part of the reduction in aggregate number is due to the fact that aggregates are motile during this stage of development, and that two aggregates may merge into a single aggregate if they collide. When this occurs it reduces the number of aggregates by one, and the merged aggregate is larger than either of its two constituent aggregates; both effects work to increase average aggregate size.

Aggregation dynamics are highly reproducible, so that the timing and peak values vary relatively little between the seven replicates used in this analysis. Panels A and C show the typical results for individual frames whereas panels B and D correspond to the average of all seven image stacks.

3.2.2. Spatial distribution of aggregates shows ordering with developmental time

There is a significant rearrangement of aggregates between an 'early' stage, when majority of aggregates are initially formed (~7h after the onset of starvation) and a 'late' stage (~24h), when a population of aggregates has stabilized. Could these rearrangements alter the spatial distribution of aggregates on a two dimensional surface? Specifically, are the aggregate distributions at early and late stages random or ordered, and does the distribution of aggregates become more or

less ordered as development progresses? To quantify any changes in aggregate order, CRDFs were calculated for aggregates in experimental images at both early and late stages, and these were compared the CRDFs for randomized distributions of the same aggregates (see 2.3.1).

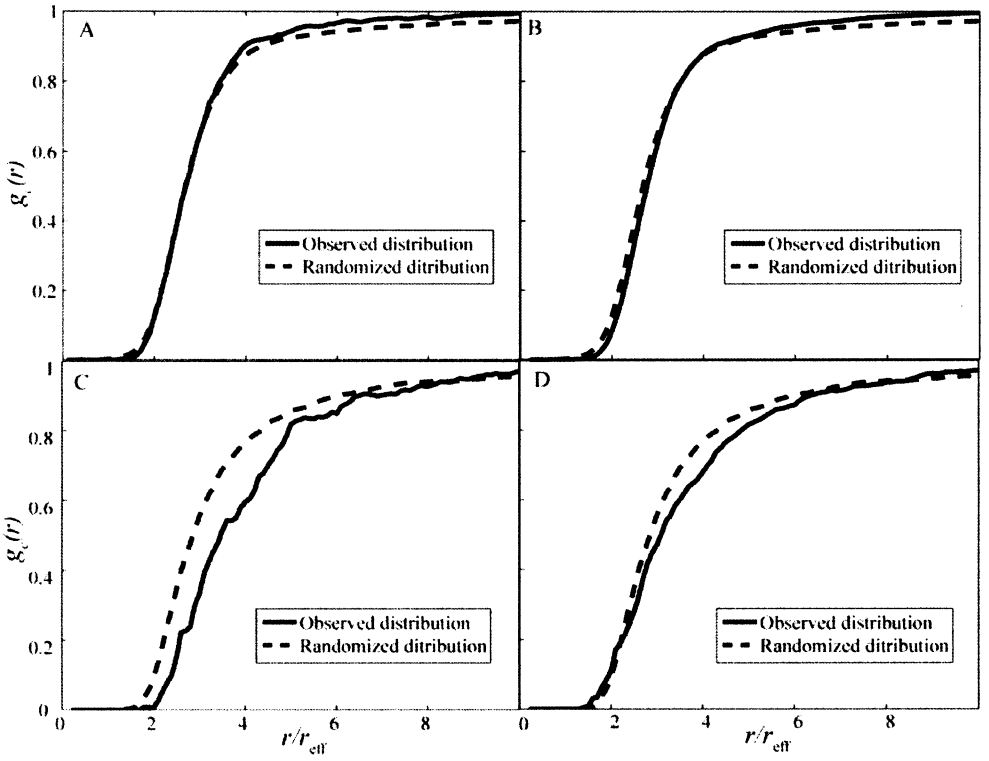


Figure 3-3 Cumulative radial distribution function is used to characterize the spatial distribution of fruiting bodies and compare it to random distribution. (A, B) Distribution at the early stage of fruiting body development (time of the maximal number of aggregates); (C,D) Distribution at the end stage of fruiting body development. Panels A and C represent the data from one representative movie; panels B and D show an averaged data from 7 different movies recorded on different days.

Figure 3-3 A and B display the CRDFs at the early stage when the number of aggregates was highest (i.e., the peak of the curve from Figure 3-2 A and B). The results indicate that distributions obtained both from an individual image stack

(Figure 3-3 A) and from an average of seven image stacks (Figure 3-2 B) are very close to the randomized controls. To statistically test if the cumulative distributions are different, the differences are computed in the area below the two curves for all the pairs of samples from the seven experimental and 30 randomized distributions, and performed a standard t-test to see if the mean of this distribution is significantly different from zero. The results indicate that CRDFs computed from images at the early stage of aggregation are not significantly different from randomized distributions (p-value>0.05). The same analysis applied to images from the late stage (i.e., when aggregates had stabilized), produces very different results; CRDFs computed from images at the late stage (i.e., the final image of the image stack (Figure 3-3 C and D) lie sufficiently below their corresponding randomized distributions that there is a statistically significant difference (p-value<1e-8).

3.2.3. Traffic jam aggregation model reproduces some but not all features of aggregation

Developmental aggregation of *M. xanthus* has been a subject of significant attention of the modeling community and multiple mathematical models of aggregation have been published[32, 34, 39, 41, 56]. Nevertheless, few of these models explicitly study dynamical aspects of aggregation, such as the merging or disappearing of aggregates, or changes in the distribution of aggregates on an agar surface. Whether a mathematical model of aggregation can reproduce the time-dependence of aggregate number and size as well as non-randomness of aggregate distribution needs to be tested. To this end a two dimensional model of aggregation

is constructed based on the "traffic-jam" idea [32, 33, 39]: cells slow down or cease motility at high cell density (see method). It should be noted that the nature of the inquiry required to use a model of aggregation on a length-scale at which multiple fruiting bodies will be formed (2x2mm). Because of these constraints it is no possible to adopt any of the more sophisticated 3D model of aggregation [34, 41], as it would not be computationally feasible.

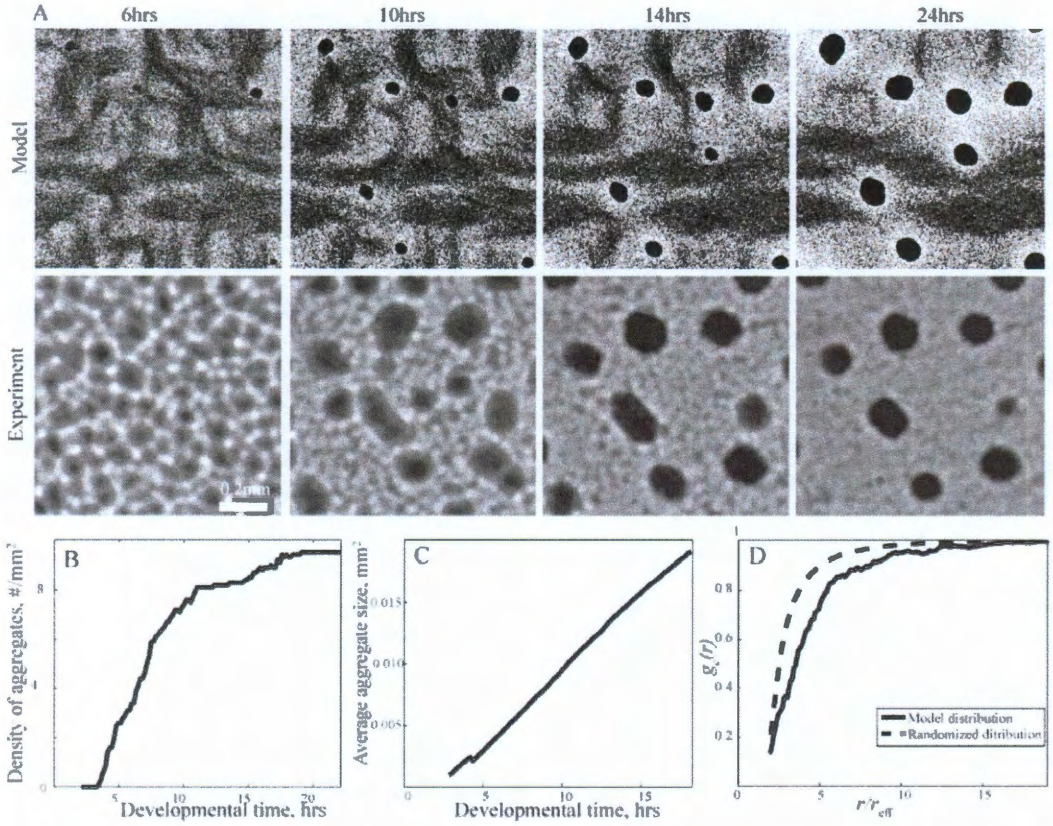


Figure 3-4 Aggregation dynamics in agent-based mathematical model. (A) Snapshots from the aggregation movies from traffic-jam model (top row) as compared to experimental snapshots at this same scale (bottom row). (B) Number of fruiting bodies in 1mm2 area monotonically increases with developmental time. (C) Average size of fruiting bodies monotonically increases with developmental time. (D) Cumulative radial distribution function of the end stage of traffic-jam model development is significantly

different from random. Results in panel B, C and D are averaged from 10 simulations with the same parameters and initial conditions.

Images of the mathematical model of aggregation (1 mm^2) are shown in Figure 3-4 A, top row. Starting with the parameter set used by Sliusarenko et al then some are adjusted to ensure the same number, density, and size of aggregates as observed in experiment (snapshots of the experimental images cropped to 1 mm^2 are shown on the bottom row for comparison). Although a comparison between model and experiment show that their endpoints appear similar, their dynamics are dissimilar and changes in aggregate number and average aggregate size as a function of time effectively illustrate the dissimilarity (Figure 3-4 B and C). In the model, aggregate number increases monotonically with time, the increase is less rapid than in experiment, and there is no plateau or subsequent decrease. Average aggregate size shows a nearly linear increase with time in the model, with no plateau between 10 and 15 hours and no gradual decrease, as is observed in experiment. There is also little or no aggregate movement or merging in the model, and very few aggregates regressed or disappeared. In fact, there is no measurement related to the dynamics of development for which the model parallels experimental results. And it must be noted that, in all simulations performed, aggregate growth rate was significantly slower in "traffic jam" model as compared to experiments and very little aggregate dispersal was observed.

Because of these differences, the early and late aggregate distributions are examined in the model for the verification of if they were similar to experiment. By calculating the model CRDFs for both early and late distributions in the same

manner as experiment, it is clear that the distribution of aggregates in the model was ordered at both early and late stages. Therefore, model CRDF data also disagree with experiment, which is random at the early stage and ordered at the late stage. Apparently, swarm development in both model and experiment finishes with ordered set of aggregates, but the path is different.

An *M. xanthus* cell on an agar substrate is capable of changing its reversal frequency and possibly gliding velocity in response to signals [51, 52, 68], but its ability to 'steer' in two dimensions on a surface appears to be extremely limited; it can turn in response to distortions within the agar (elasticotaxis) [132], and it can bend in response to collisions with objects such as other cells [133]. This cell movement is observed in both isolated cells and individual cells within a swarm, but it is clearly insufficient to explain how a randomly distributed dense population of millions of starving bacteria can synchronously break symmetry and self-organize into an ordered field of aggregates [30]. Much of the mechanics underlying this process has yet to be described.

3.3. Discussion

Multiple mathematical models with various modeling formalisms, parameters and levels of complexity that result in aggregation have been created [27, 34, 39, 41, 134]. However, for a simulation to accurately represent the collective behavior of cells within a developing swarm, it must accurately capture the observed aggregation dynamics on a variety of scales, from a single cell to a single

aggregate to a multiplicity of aggregates. The goal of this work was to quantify features that describe aggregation over the largest scale, which is the entire swarm.

A comparison of the traffic-jam model to experimental data shows that this model can successfully reproduce end-state aggregation, but not the dynamics that lead to that end-state. Through parameter optimization a simulation that reproduced the correct aggregate size as well as the correct number of aggregates per unit area can be generated. The simulation also reproduced an ordered aggregate distribution that was similar to experimental observations at the end of development. On the other hand, the simulation was unable to reproduce many important features of aggregation dynamics. Simulated aggregates formed much less synchronously, exhibited a steady increase in size until the end of the simulation, and did not exhibit significant aggregate disappearance. As a result, the distribution of *in silico* aggregates never exhibited an increase in ordering with time. All attempts to resolve this disagreement through parameter optimization were unsuccessful. Because of these discrepancies between simulation and experiment, the traffic-jam model alone does not sufficiently explain the dynamics of aggregation or elucidate the mechanistic processes that result in fruiting body formation.

Based on the analysis of experimental data, there are several quantifiable swarm behaviors that must be represented in a development simulation in order for it to be considered accurate. Aggregation must be rapid and synchronous. Early aggregates must appear within the first few hours, aggregate number must increase rapidly for no more than two hours, and thereafter must rapidly decrease by more

than half over the next three hours. Aggregates must not all share the same fate during the period of decreasing aggregate number; some must grow, others must regress, and still others must move and/or merge with other aggregates. During this period, average aggregate size must steadily increase until it reaches a plateau between 10 and 15 hours, at which point these late aggregates must appear stable and exhibit an ordered distribution. This distribution must be more ordered than the distribution of early aggregates, which must be closer to random. In other words, ordering must be imposed on the population during the period between early and late aggregation.

Having stated this set of swarm behaviors, it should be noted that some of these may be sensitive to differences in experimental methods. As the description of swarm dynamics becomes more quantified and detailed, it becomes increasingly likely that at least some behaviors may occur only under a limited set of conditions. Variation in parameters, such as substrate agar concentration, degree of starvation stress, or initial population size and density, will result in somewhat different swarm behaviors. The quantified swarm behavior that occurs under a specific set of relatively standard experimental conditions, but it is not yet known which of these behaviors will prove to be robust. Nevertheless, any mathematical model that aims to explain development in *M. xanthus* swarms must be able to reproduce these behaviors using some set of optimized parameters.

Chapter 4

Statistical image analysis reveals features affecting fates of *Myxococcus xanthus* developmental aggregates

4.1. Introduction

Despite extensive use of microcinematography to examine aggregation [37], interpretation has been mainly qualitative. A quantitative metric is essential to assess the agreement of experimentally observed aggregation patterns to those produced by mathematical models aiming to reproduce the morphogenesis *in silico*. While various groups identified overlapping but distinct sets of model ingredients that lead to aggregation in computational simulations [27, 32, 33, 39, 135], further quantification of experimental data is essential to refine these models.

Here statistical image analysis and feature extraction methods are used to quantitatively characterize time-lapse images containing thousands of nascent

aggregates. Aggregates are automatically tracked in space and time to identify their fates. A list of 33 parameters (features) is prepared that characterize each nascent aggregate and cluster these features into 4 major classes. Thereafter statistical image analysis is used to identify the main features controlling aggregate fate during dispersal, fusion and splitting.

4.2. Results

4.2.1. Dynamics of aggregation as depicted by microcinematography

The most widely used laboratory conditions for initiating *M. xanthus* development were employed. Under these laboratory conditions the pattern of fruiting bodies is so dense that many aggregates are in close proximity with one another with about 160 fruiting bodies in the field of view (Figure 4-1). Within the nascent fruiting body the majority of the cells undergo Programmed Cell Death (PCD) while the remainders form myxospores.

This work focused on the dynamics of aggregate formation and fate. Initially many more aggregates form than survive to become fruiting bodies (Figure 4-1 A). Over a period of about 24 hours many of the recede into the cell mat where the cells disperse to join other aggregates (Figure 4-1 B). In some cases aggregates in close proximity merge to form one large fruiting body (Figure 4-1 C). In rare cases unusually large aggregates split to form two smaller fruiting bodies (Figure 4-1 D).

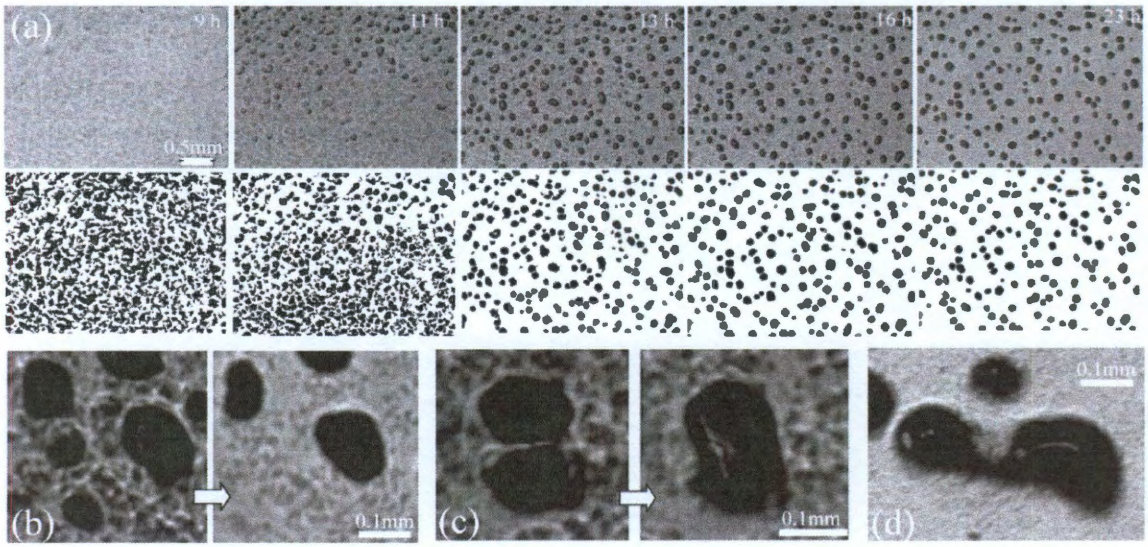


Figure 4-1 : Image sequence describing *M. xanthus* aggregation dynamics (a) Selected frames from time-lapse microcinematography. Times are noted in the upper right hand corner of the frame. Aggregation begins at approximately 9 h and the number of aggregates decreases beginning about 13 hours. Bottom row shows binary images processed for aggregate detection (cf. Supplemental methods). (b-d) Various fates of transient aggregates during development: (b) aggregates stay or disperse; (c) two aggregates merge; (d) one aggregate splits into two.

4.2.2. Image Segmentation, Feature Extraction, and Quantification

For this analysis, properties of transient and stationary aggregates are emphasized. The first step involved separating the aggregates from each time lapse image into individual aggregates and surrounding spaces using a combination of standard segmentation algorithms based on Markov-random field (MRF) [120](see 2.2.1.2 and 2.2.2 for details). Segmentation is based on the idea that aggregates are darker in intensity than the surrounding field. A label is applied for each pixel, label 1 for pixels within aggregates and label 0 for the remaining pixels. Assuming labels vary smoothly everywhere except at aggregate boundaries and that the gray scale

image intensity of the original image is related to segmentation, segmentation is reduced to energy function optimization including terms penalizing discontinuity and rewarding correlations between gray scale and binary intensity. A lighting intensity correction is applied to compensate for possible uneven illumination of images. The segmented images are thereafter converted to binary images with black aggregates surrounded by the white spacing (Figure 4-1 (a), bottom row).

Segmentation is applied to each timeframe image then linked with subsequent images by detecting overlapping aggregates. This approach allows to discern the three aggregate fates depicted in Figure 4-1 B-D. Statistical analysis of the images was used to decrypt properties of individual aggregates that influence their fate. To this end an extensive list of features (parameters) was devised that can characterize each aggregate according to its geometry, proximity to other aggregates, and size. A complete list of the 33 features is listed in Table 4. The representative features include distance to nearest neighbor, aggregate area and perimeter, area of the nearest neighbor, total/average/minimal/maximal area of surrounding neighbors, solidity, eccentricity, and others. The list includes features defined as ratios of individual features. All 33 features are automatically identified for each aggregate present in each frame of the time-lapse movie. The list of the features introduced is quite exhaustive and sufficient for purposes but can nevertheless be further extended with new features or mathematical functions of existing features.

The features identified in the list are not independent and may show a high degree of correlation. To further identify biologically meaningful features of aggregates the list of 33 features is automatically identified for ~160 aggregates from the last time-frame of the aggregation movie and use these data to compose 33 vectors (one for each feature) containing the values of these feature for each of the aggregates. The Spearman Correlation between the features is calculated and Euclidean distance between correlation coefficients is used to build a binary tree-based clustering of the features. In the end, features that cluster together show the highest correlation. The features divide into four major clusters representing features associated with aggregate's proximity to neighbors (#1-3), various size parameters (#4-15), parameters of the aggregate's neighbors (#17-28), and image shape and topology (#30-33) (Figure 4-2). It is interesting to note that aggregate orientation (#29) clustered together with shape-related features. On the other hand feature #16, the ratio of the nearest-neighbor distance (feature #3, belongs to the proximity cluster) to the equivalent diameter (feature #14, size cluster) clusters separately.

4.2.3. Small aggregate size but not neighbor-related parameters correlates with aggregate dispersal

About 50% of the aggregates formed by 13 hours develop into mature fruiting bodies with the majority simply dispersing into the cell mat. To determine which of the features (or feature classes) have the largest influence on the probability of aggregate dispersal, Mutual Information (MI), the quantity that

measures the mutual dependence between two random variables, is computed (**Error! Reference source not found.**and **Error! Reference source not found.**).

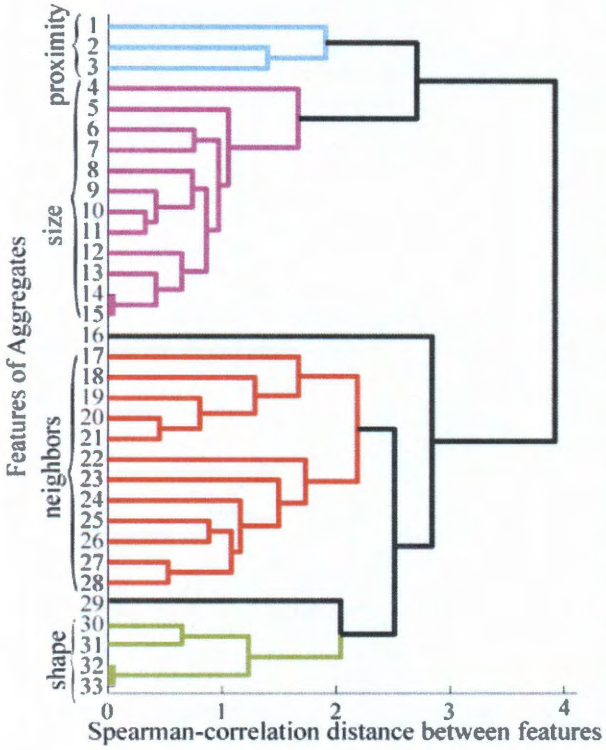


Figure 4-2 Characterization of *M. xanthus* aggregate features derived from microscope images. A set of 33 features encompassing multiple aspects of each aggregate was automatically detected for more than 150 aggregates from the last frame of time-lapse movies. This data was used to compute spearman correlation based distance between features and cluster these features. Four of the clusters have biological relevance: features associated with aggregate proximity to its neighbors (blue); features related to aggregate size (purple); features associated with neighbors of a given aggregate (red); and features associated with aggregate shape (green). Feature #16 (black) corresponds to the ratio of two features from different clusters (the distance to nearest neighbor over aggregate diameter) and clusters separately from either.

Normalized mutual information for all 33 features unexpectedly revealed that only features associated with size show high mutual information with dispersal (Figure 4-3 A). Aggregate area and equivalent diameter show the highest values. Neither proximity to other aggregates nor parameters of their neighbors (such as size) affect aggregate dispersal. Distribution of areas of vanishing or stable aggregates (Figure 4-3 a, inset) shows a clear separation between the sizes of these

aggregates with approximate threshold of 6000 μm2 separating aggregates that are very likely to disperse from those likely to mature into fruiting bodies.

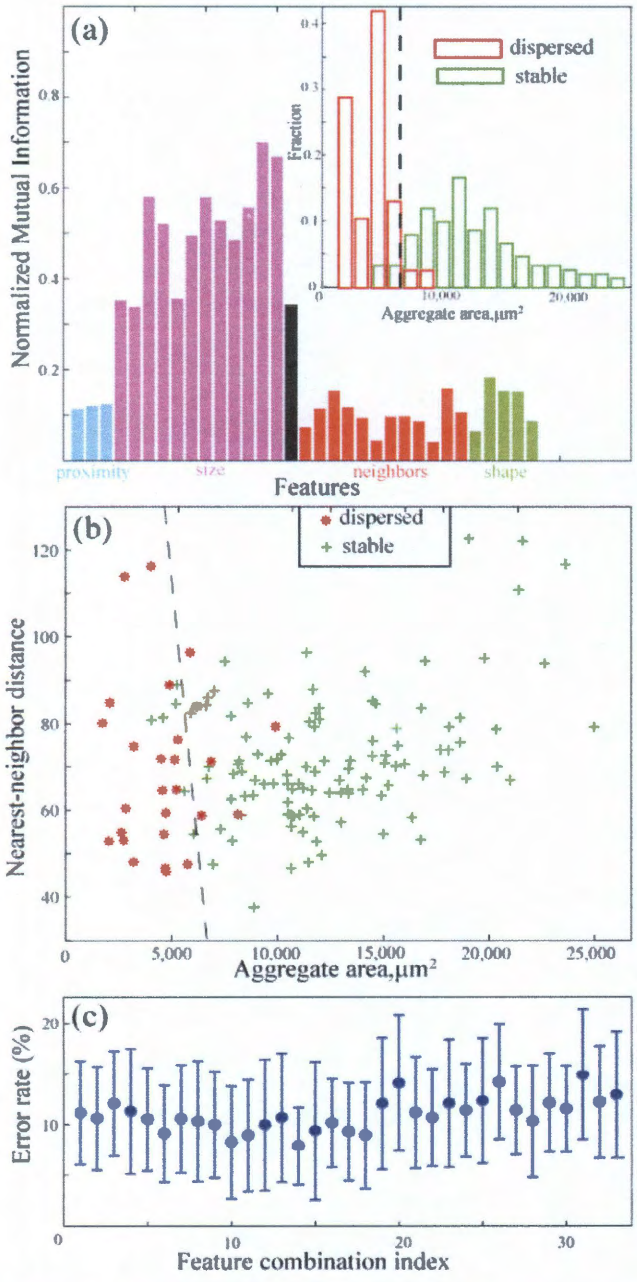


Figure 4-3 Features correlated with aggregate dispersal. (a) Only size related features are significantly correlated with aggregate dispersal. The inset shows a histogram of aggregate areas at intermediate frame corresponding to 13.5h for aggregates that remain stable until the end of the movie (green) or disperse (red). The dashed line depicts the threshold area that predicts the aggregate fate. (b) Distance to a neighboring aggregate does not significantly affect aggregate stability. The support vector machine boundary (dashed line) separating stable (green +) or dispersing (red *) aggregates runs nearly parallel to the distance axis. (c) Single size-related features serve as an accurate (~10% error) predictor of aggregate dispersal. Support vector machine analysis shows that combining additional features does not improve prediction accuracy. Each point corresponds to error rate obtained from combining *n* features with highest mutual information (panel a).

To determine whether features in addition to size (area) increase the ability to predict aggregate fate, pair wise combinations of features were considered,

aggregate area and another feature from a different cluster. Figure 4-3 b shows a representative scattered plot using a combination of aggregate area with nearest-neighbor distance (NND). This data was used to train a support-vector machine to determine linear boundaries separating stable and vanishing aggregates in this plane. The resulting line runs nearly parallel to the y-axis indicating that NND has little effect on aggregate dispersal. The data was separated into two unequal parts. 75% of the data was used to train SVM and the remaining 25% to estimate the error rate of such predictions. The error rate is defined as the percentage of cases (both false positive and false negative) where dispersal is not correctly predicted by a threshold model. The error rate for a 2d SVM model (Figure 4-3 b) is 10.8±6.16% and is not significantly different from that predicted by the area alone (11.7±4.95%). Similar results are obtained when other neighbor-related, proximity-related or shape-related features are combined with aggregate area. As a result, aggregate size is the only feature significantly affecting its dispersal.

To examine combinations of more than two features, the features are placed in order of decreasing mutual information with dispersal (data of Figure 4-3 a) and then combined, top n features for n = 1, ...33. For each combination SVM is used to predict the hyper-plane boundary between stable and dispersing aggregates and the error rate for each predicted boundary is computed. The results indicate that combining 2 or more features does not increase prediction accuracy.

4.2.4. Close proximity is necessary but not sufficient for aggregate merging

To determine which features enhance the ability of an aggregate to merge with its neighbor, expressions similar to those in **Error! Reference source not found.** and **Error! Reference source not found.** were used to compute the normalized mutual information between aggregates merging and the list of 33 identified features (Figure 4-4 a). Only proximity-related features such as distance to the nearest neighbor have significant codependence with merging. Other factors, such aggregate size or nearest-neighbor size show little predictive power. The inset shows a normalized distribution of nearest-neighbor distances for aggregates that merge and those that do not. The results indicate that merging can only happen if aggregate separation (measured at 13.5 h of development) is smaller than a certain threshold (~150 μm). However, only 28% of aggregates with centers spaced closer than this threshold will merge and the remaining 72% remain separate. Thus nearest neighbor proximity is required for merging but is not sufficient.

4.2.5. Aggregate size and eccentricity affect its ability to split

In relatively infrequent instances a single aggregate at 13 hours of development later splits into two smaller ones (Figure 4-1 d). To ask which features are associated with aggregate splitting normalized mutual information (see section 2.2.2.2) was computed between splitting and aggregate features. As shown in Figure 4-4 b both aggregate size (e.g. area) and shape (e.g. eccentricity -- ratio of minor to major axis of the corresponding ellipse) show significant codependence with

splitting. The inset shows that splitting occurs only with eccentric aggregates (eccentricity above ~0.8). Splitting is also more likely in the aggregates with a large area (data not shown). It is expected that combinations of eccentricity and aggregate size will be an even better predictor of aggregate splitting but given the rareness of splitting insufficient data to test this hypothesis is available.

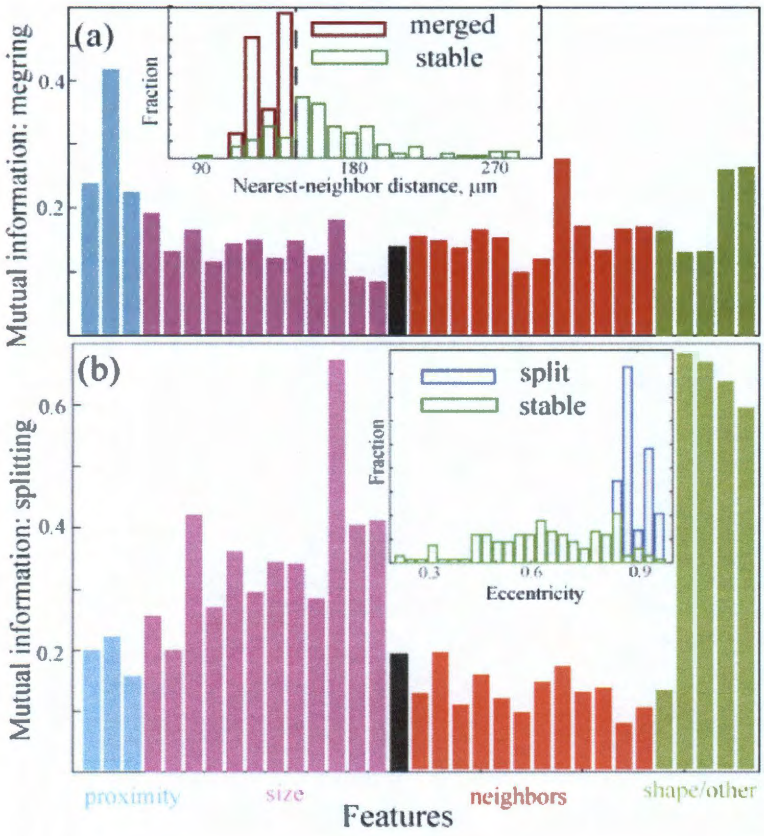


Figure 4-4 Features affecting aggregate merging and splitting. (a) Distance-related features have the highest mutual information regarding whether an aggregate merges with another. The inset shows a histogram of nearest-neighbor distance at 13.5 h for aggregates that merge (purple) or do not merge (green) by the end of the movie. Close proximity to another aggregate is necessary but not sufficient for merging. (b) Aggregate splitting is affected by the aggregate shape/eccentricity and size. Most aggregates have a circular base. Oblong aggregates often resolve into smaller ones. Aggregate perimeter (feature #13 clustering with size) also shows high mutual information with splitting and is large for eccentric aggregates. The inset shows a histogram of aggregate eccentricity measured at 13.5 h for aggregates that split (purple) or do not split (green) by the end of the movie. High eccentricity of an aggregate is a good predictor of splitting.

4.3. Discussion

4.3.1. Statistical image analysis to discover aggregation metrics

Time-lapse microcinematography of *M. xanthus* wild type aggregation revealed three unusual aggregate fates, dispersal, fusion, and splitting. Quantitative analysis of this data is expected to predict features correlated with each type of fate. Here a quantitative metric is developed to describe each aggregate and then determine the features that affect the fate of each aggregate. The metric consists of 33 features that measure various morphological and spatial parameters of the aggregate as well as its neighbors. Some of these features are mathematically interdependent, like aggregate's area (A, feature #15) and aggregate effective diameter ($D = 2\sqrt{A/\pi}$, feature #14). Other features are strongly correlated in nonlinear fashion, e.g. average area of neighbors (feature #28) and average effective diameter of neighbors (feature #27). Therefore, hierarchical clustering based on non-linear (rank) separates features into groups. The clustering resulted in four biologically meaningful groups of features – those associated with properties of the aggregate itself – shape and size, and those associated with the aggregate's surroundings – proximity to neighbors and neighbor-size related features. Even though the metrics proposed can be extended, the four classes considered here distinguish between biologically-distinct interactions mechanisms.

The features considered in this study neglect two potentially important properties of aggregates because preliminary analysis did not identify benefits for including either. First, the analysis is restricted to static features, i.e. features based

on a single time frame. Dynamic features, i.e. features computed based on multiple time frames such as rate of change of features can provide additional metrics. Moreover, the analysis is limited to features that can be extracted from binary (two-color) images of aggregates and therefore ignore features that take into account grey-scale intensity of the individual aggregates.

4.3.2. Aggregate disappearance is independent of long-distance signaling

Despite decades of research, the biophysical mechanisms that mediate *M. xanthus* aggregation remain a mystery. Even though the *M. xanthus* genome contains 8 chemosensory gene clusters [5], and one of these mediates chemotaxis to certain lipids [136-139] conclusive proof that cells follow an attractant gradient during aggregation is lacking. Mathematical modeling has been an important tool in identifying the "minimal set" of interactions that can lead to aggregation in silico [27, 34, 39, 41]. However, comparison between modeling and experimental aggregation has been primarily qualitative. Moreover, existing models pay relatively little attention to the two-dimensional distribution of aggregates or the fates of transient aggregates such as splitting, merging and dispersal. For instance, the traffic jam model of aggregation in which aggregation is driven by cell density-dependent motility inhibition grossly underestimates aggregate disappearance[34].

Three possible scenarios determining whether aggregates mature into a fruiting body or disperse can be envisioned. First, in a chemotactic model for aggregation cells from smaller aggregates might join larger aggregates due to a stronger attractant gradient. Second, cells could leave and join aggregates in random

fashion but stay longer in the larger aggregates [32]. As a result larger aggregates will grow at the expense of their smaller neighbors analogous to Oswald ripening in solutions[140]. Third, insufficient cell influx due to closely spaced aggregation foci competing for a limited number of cells may lead to aggregate destabilization [41]. Importantly, all three scenarios relate aggregate dispersal to factors extrinsic to the aggregate itself, i.e. size and proximity of the neighbors. The analysis presented here (Figure 4-3) shows that proximity with other aggregates is not the major feature aborting their development thereby disproving all of the above scenarios. The only intrinsic aggregate property that initiates dispersal is small size.

The reason development bogs down in small aggregates is a matter of speculation. One possibility is that the small aggregates fail to differentiate the necessary cell types to progress further, namely those able to undergo PCD and sporulation. Developing cells produce a third cell type, the peripheral rods, which show little ability to aggregate [141] and express early developmental genes but not late developmental genes [25, 135]. Cross sections of fruiting bodies show that peripheral rods are found only in the lower layers [142]. Perhaps aggregates that abort development contain only peripheral rods or vegetative cells that are unable to differentiate further.

A second reason aggregates disperse could be the absence of an essential morphogenic signal. Short-range, quorum-sensing signals within the aggregate could explain how the transient size of an aggregate behaves independent of neighboring aggregates. Such signaling could be facilitated by direct cell contact, by

a slowly diffusing signal, or one with a very short half-life. The diffusion constant can estimated from an average distance between aggregates ~150 μm over the time course of aggregate disappearance of roughly 3 hours to be less than 2 μm2/s. There are two candidates for such signals, C-signal and E-signal. C-signal mutants fail to aggregate; whenever aggregates of any size are produced the aggregates disperse [37]. These results suggest that the C-signal is required to stabilize growing aggregates. C-signal increases throughout development [49, 54]. Since C-signaling is enhanced by end to end contact [48], it may be especially high in fruiting bodies. In contrast E-signal mutants aggregate suggesting that E-signal functions to destabilize aggregation. E-signal is likely to be a lipid containing the fatty acid iso 15:0 and produced in lipid bodies, the major metabolic product of early development [142]. Though the mechanism of E-signaling remains unknown, lipids would not be expected to diffuse very far and could serve as a short-range signal. Lipid body synthesis increases throughout the aggregation period and occurs only within fruiting bodies [142]. The stability of aggregates could be determined by the relative ratio of both signals.

Analysis of merging aggregates provides further evidence against the presence of chemoattractant gradients guiding cells into aggregates. If long range, horizontal gradients of a chemoattractant emanate from a fruiting body stronger gradients are expected around larger aggregates and therefore size of the aggregate or size of its nearest neighbor would be significant factors in merging. This expectation differs from the results in Figure 4-4 a, where the spacing between aggregates is the only significant factor affecting their merging.

Chapter 5

The Mechanism and Physiological Role of *Myxococcus xanthus* Predatory Rippling Behavior

5.1. Introduction

M. xanthus cells employ two genetically distinct forms of motility to move on solid surfaces by periodically switching their gliding direction by polar reversal [43, 59]. These bacteria have been well studied for their ability to execute a complex multicellular developmental program that includes aggregation into multicellular mounds, termed fruiting bodies, and differentiation into dormant, environmentally resistant myxospores [88]. *M. xanthus* is also known for its cooperative predation of other microorganisms by collectively spreading towards prey cells with surface motility and by producing antibiotics and lytic compounds that kill and lyse prey cells [10, 143]. One of the most puzzling forms of collective dynamics exhibited by

M. xanthus is their ability to self-organize into ripples -- bands of travelling high density waves that dictate biofilm behavior [6, 8, 9, 78]. Although the *M. xanthus* counter-traveling waves appear to pass through each another, they actually reflect off of one another and are termed "accordion waves" [6-8, 31, 36]. These waves are distinct from the waves originating from Turing instability diffusion-reaction patterns, such as those in chemical systems or observed during development of the other well-studied model social microorganism, the amoeba *Dictyostelium discoideum[144-146]*.

Until recently, the mechanisms underlying *M. xanthus* ripples were studied exclusively during starvation-induced development. However, more robust rippling during predation was demonstrated recently by the Kirby laboratory. The application of mathematical modeling to uncover the mechanism of developmental ripples revealed that the wave properties are consistent with contact-induced reversal signaling [7, 8, 31, 36]. This signaling was hypothesized to originate from 'head-to-head' collisions of cells moving in opposite directions and to resulted in an exchange of C-signal that accelerated the reversal clock [6, 7, 36]. C-signal is an extracellular morphogen that controls developmental timing by sequentially inducing aggregation and sporulation via its contact-dependent pole-to-pole transmission. Developmental aggregation and alignment are induced through the C-signal-dependent stimulation of the *frz* chemotaxis-like system, which includes an unconventional soluble cytoplasmic chemoreceptor homologue FrzCD. Recently the Zusman laboratory has shown that FrzCD appears to transiently align and stimulate reversals in growing cells making side-to-side contact [59].

With the discovery and characterization of predatory rippling behavior, questions concerning the underlying mechanism and physiological role of the ripples reemerged [9, 10]. Is the mechanism of predatory ripples similar to that of the developmental ripples? And what is the physiological role of this self-organization behavior in predation? Investigate was performed for these questions with a combination of mathematical modeling and experimentation. A mathematical model of the travelling wave is constructed by adapting the recently-proposed side-to-side reversal-induced signaling model [59] and demonstrate that robust ripples result from these interactions. This thesis also shows that the formation of ripples allows *M. xanthus* cells to cover their prey faster and to remain in the prey region for longer. Moreover, the predictions have been experimentally tested and confirmed from the model and the quantitative properties intrinsic to predatory ripples.

5.2. Results

5.2.1. A new agent-based model robustly reproduces rippling self-organization

To model predatory ripples a modeling formalism is needed that would allow to connect the motility of individual cells, inter-cellular interactions and rippling patterns. Here an agent-based model (ABM) ([36, 147, 148]) formalism is chosen that represents cells as individual agents that move and interact according to pre-specified equations. Unlike continuous, cell-density-based approaches, the ABM approach allows cell variability and modular implementation of interactions to be

easily incorporate into the model. The details and equations describing the ABM are summarized in the 2.1.2. Here is a qualitatively description of the main model ingredients that result in rippling.

Each agent in the ABM represents one cell as a self-propelled rod on a 2-D surface. The rods move continually along their long axis with periodic reversals by switching polarity of the two ends. As in the previous models [20-23], it is expected that the ripples to emerge from the synchronized cellular reversals. A plausible mechanism for such signaling can be extracted from the observations of Mauriello et al. [14], in which as two growing cells pass by one another making transient side-to-side contact, their FrzCD clusters align and one or both of the cells reverse. The reversals stimulated by this intercellular signaling would be similar to reversal-inducing head-to-head collision signaling hypothesized to occur due to C-signal exchange during *M. xanthus* development. Taking this into account, the following three ingredients are necessary for reversals in response to cell-cell signaling:

i. Two countermoving cells that make side-by-side contact with a length overlap of 50% or more (Figure 2-4) have a probability of engaging in a signaling event that results in reversal of at least one of the cells.

ii. Cells have a minimal reversal time, so there is a refractory period after each reversal during which another reversal/signaling is not possible. To be consistent with the observed reversals of isolated cells, the possibility of spontaneous reversals is also introduced without signaling with a native period three times the refractory period.

iii. The cells must align locally.

It appears that these three ingredients are sufficient for an ABM to produce rippling behavior (Figure 5-1, top row). Starting from a uniform distributions of aligned cells the model results in a population of cells that self-organize into periodic traveling bands, termed ripples, within about 3 h. The side-by-side signaling mechanism appears to be more robust than the previously postulated head-to-head signaling (Figure 5-2). Moreover, the existence of the ripples seems to critically depend on the probability of reversal-inducing signaling events (Appendix figure 1 E). If this probability is too small (less than 5%), then no ripples will form and the cells will remain uniformly distributed on the 2-D surface. Once the signaling probability exceeds a specific threshold (~ 5--10 % for the parameters used here) the uniform distribution is unstable and cells self-organize into ripples. This observation led to hypothesis that the presence of prey cells or some component released upon their lysis activates side-by-side signaling and thereby results in the formation of ripples in the regions where prey is present. Note that a small value for the probability of signaling is consistent with the observation that in most cases only one of the cells engaged in side-to-side signaling reverses [14]. If the probability of signaling is low and signaling in each direction is independent, it is unlikely that these signaling events will occur simultaneously. Furthermore, once one of the cells reverses, both will be moving in the same direction and as a result

no-longer signaling to one-another.

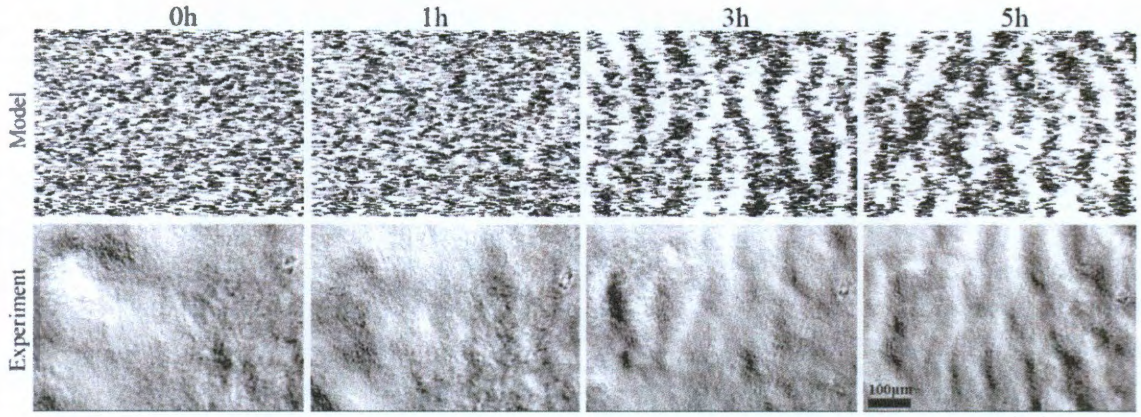


Figure 5-1 Comparison of ripple initiation in the ABM simulations (top row) and experiments (bottom row). The timing of the snapshot for each row (0, 1, 3 and 5 hrs) is indicated for each column. Initial time (0 hr) corresponds to initiation of the ABM simulation with a uniform cell distribution or the time *M. xanthus* cells fully cover the prey in the field of view. After about 3 h the cells self-organize into ripples in both the ABM simulations and the experiment. The ABM simulations qualitatively match the experimental rippling pattern as the wavelength and crest shapes are very similar. The fields of view of both the ABM simulation images and experimental images have the same dimension that is shown by the black scale bar of 100 μm.

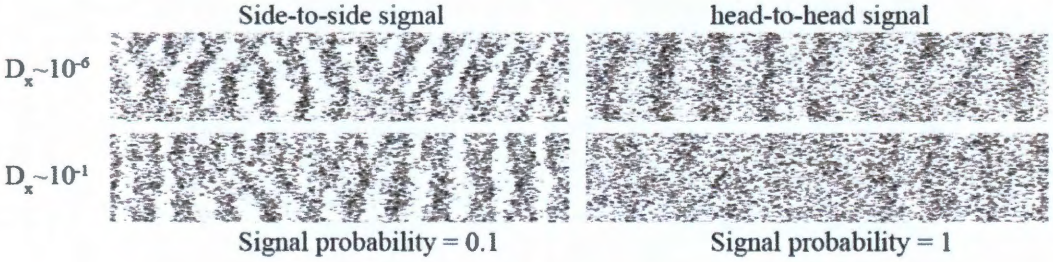


Figure 5-2 The new side-to-side signaling mechanism is compared to the previous head-to-head signaling mechanism. Although both mechanisms can produce ripples at a low noise level the side-to-side signaling mechanism is much more robust. To produce ripples in the ABM the head-to-head signaling mechanism must have 100% signal probability, whereas the side-to-side signaling mechanism only needs 10% signal probability, which is a more realistic value. When the noise level is increased to match the value obtained in the experiments (about 25% of the mean), only side-to-side signal can

produce waves (bottom row); the head-to-head signal does not produce visible ripples even with 100% signal probability. High noise level is obersved in the experimental cell movement measurements. Conclusion is that the side-to-side signaling mechanism is more realistic.

5.2.2. Quantifying individual and collective cell behavior in predatory ripples with fluorescent microscopy

To test the modeling predictions experimentally, a protocol is designed to monitor cell motility on a solid nutrient agar surface in the presence of prey cells. Ripples were observed with fluorescence and differential interference contrast (DIC) time-lapse microscopy allowing to track individual fluorescently-labeled cells as a small percentage (~0.1%) of an unlabeled wild-type population in the presence of differentially labeled prey. As a result information is obtained concerning both the global properties of the ripples, such as wavelength (distance from one wave crest to the next), wave crest width and at the same time collected behavioral measurements of individual cells including their coordinates, velocity, reversal period, position and time of their reversals

These data provided crucial input into the model and allowed for tests of modeling predictions. It is noteworthy that the experimental ripples appear very similar to those produced in the simulation (Figure 5-1, bottom vs. top row). To compare the timing of wave initiation between the mathematical model and the experimental results, the time point when *M. xanthus* cells fully cover the prey in the field of view was chosen as the starting time (0 h in Figure 5-1 bottom row). Snapshot images at 1 hr, 3 hr and 5 hr were selected to show the process of ripple formation in both systems. The experimental process of wave initiation appears to

follow the same dynamics as the simulations. These results indicate that the ABM is capable of qualitatively reproducing ripple pattern dynamics as observed under experimental conditions. Interestingly, waves generated with ABM appear somewhat more focused than those the experimentally observed ripples which have a smaller density gradient from crests to troughs. This observation suggests that perhaps not all the cells in the biofilm participate in rippling behavior.

5.2.3. Wavelength quantification is consistent with proposed rippling mechanism

To quantitatively compare the rippling patterns produced by the mathematical model to those of the experiments, a method is needed to quantitatively characterize the ripples and relate their patterns to the behavior of individual cells. For developmental ripples, theoretical analysis [22, 23] proposed a simple expression relating the wavelength of ripples (λ) with individual cell speed (v) and cell reversal periods (τ):

$$\lambda = 2v\tau \qquad \text{Eq. (16)}$$

Biologically this means that cells in two oppositely moving crests (relative speed) reverse their directions every time the crests are superimposed. This prediction was in agreement with experimental measurements and the results of the ABM [21]. However, the measurements of Berleman et al. [15] on predatory waves showed a slightly larger wavelength compared to the prediction of Eq. (16).

Indeed, in experiments with a measured v=3 μm/min and τ=8 min [15], their observed wavelength was about 70 μm, despite of the calculated value of 48 μm.

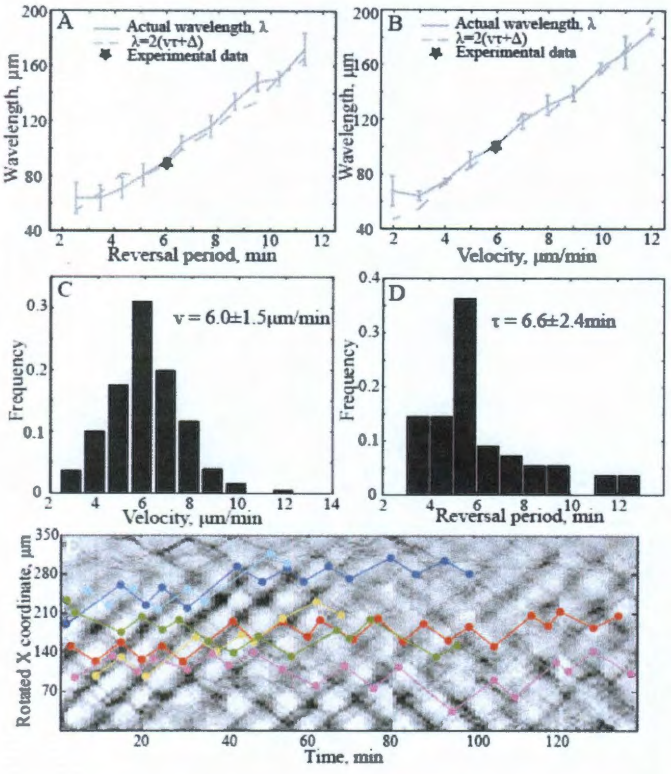


Figure 5-3 The relationship between the wavelength and individual cell motility. The ABM simulations show that the wavelength linearly scales with (A) a varying average reversal period and (B) a varying cell velocity. The velocity predicted by Eq. (17) is shown by a dashed line and is in good agreement with the wavelength measures from ABM simulations. The experimentally measured wavelength (black stars) also agrees with ABM predictions based on the average velocity and reversal period values computed from the measured distributions in panels C and D, respectively. The distributions are obtained from tracking fluorescently-labelled cells in microscopic images. (E) As the ABM predicts most cell reversals occur when two wave-crests collide. Superposing the trajectories of 6 cells on a space-time plot of 1-D averaged DIC images of their corresponding ripples confirms this prediction experimentally.

To determine if the wavelength relationship (Eq. (16)) holds in the new ABM for predatory ripples two sets of simulations were conducted. First, cell speed was kept constant at 6 μm/min, while the native reversal period was varied between 5 min and 30 min (corresponding to variation between 3 min and 12 min of an actual average reversal time due to early reversals triggered by side-to-side signaling, see

Figure 5-3A, solid line). Second, the reversal period was kept constant at around 6.6 min and cell speed was varied between 2 μm/min and 12 μm/min (see Figure 5-3 B, solid line). These values correspond to the experimental cell motility parameters (Figure 5-3 C and D text below). As shown in Figure 5-3 A and B, the wavelength (λ) values calculated using continuous 1-D wavelet transform of the DIC images (see 2.3.5) still scales linearly with cell speed/velocity (v) and average reversal period (τ). However, when no-intercept linear regression of the form of Eq. (16) is used to assess the data in Figure 5-3 A and B, the regression coefficients of 15.2 μm/min for Figure 5-3 A of and for 16.1 min Figure 5-3 B are obtained. Both of these values are slightly larger than the predicted coefficients of $2v$ (12 μm/min) and 2τ (13.2min) respectively. Therefore, a slight modification of the predicted wavelength equation is necessary. Indeed, when tracking the reversal points of individual cells in the simulations, it is clear that cells in each crest initiated their reversal as soon as the leading edge of each crest met (Figure 5-4). This can be interpreted to indicate that as the cells at the front of each crest reverse, they immediately start signaling to the other cells in their crests, leading to a "chain-reaction" of signaling and reversals. This scenario would correspond to a modified wavelength equation:

$$\lambda = 2(v\tau + \Delta) \qquad \text{Eq. (17)}$$

where Δ denotes the wave-crest width. To test the modified expression, an algorithm is developed to automatically compute the average wave-[redacted] from the simulation results (see 2.3.5) and the wavelength predicted by Eq. (17) (Figure 5-3 A and B, dash line). The results demonstrate a good agreement between

measured and predicted wavelengths. Therefore, the cells in each crest only move an average distance of $\lambda-2\Delta$ before reversing again, which results in the average reversal period $\tau= (\lambda- 2\Delta)/2v$ which is equivalent to Eq. (17). The effect of this modification is expected to be larger at higher cell densities where the wave-crest width would be larger.

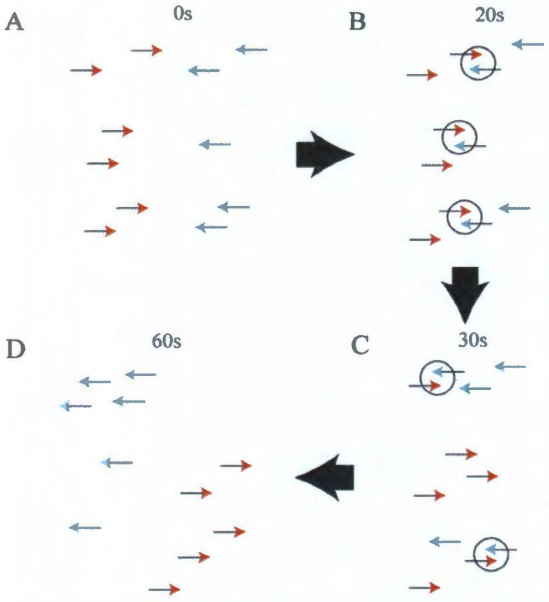


Figure 5-4 An illustration of simulation images indicating how individual cells behave and ripples are formed by as cells reverse their direction during crest edge collisions. (A) Two opposing waves approaching each other start to touch. Cells travelling to the right are marked by red and cells travelling to the left are marked by blue. (B) The initial stage of the collision of the two wave crests. Three pairs of the signaling cells are found and circled. As a result of the signaling, some cells reverse and others continue without changing the direction as it shows in (C). (C) Another two signaling events which occur between reversed cells and their previous followers in the same crests. (D) The two waves have finished collision and reversed their directions

To test the prediction of Eq. (17) on predatory waves under experimental conditions, 37 GFP-labeled individual cells were tracked within ripples for about 2 h (or until the cells left the field of view). The resulting distribution of their velocity and reversal period time-series is shown in Figure 5-3 C-E. Moreover, continuous 1-D wavelet transform of the DIC ripple images (see 2.3.5) was used to compute the wavelength and wave crest width by fitting Gaussian function to the wave crests. Using the average speed and reversal period from the histograms (Figure 5-3 C-E)

and the experimentally determined wavelength, an excellent agreement is obtained between the experimental wavelength and the simulation results (black stars in Figure 5-3 A, B). The prediction of Eq. (17) is also consistent with the data Berleman et al [15], for which v=3 μm/min and τ=8 min, Δ~10-15 μm, and the wavelength, λ, is ~70-80 μm in good agreement with Eq. (17).

According to the ABM assumptions and predictions, most of the rippling cells should travel with the wave crest and reverse, essentially as a group, as the leading edges of the two opposing wave crests collide. To test these predictions with experimental results, a method is created to view the reversals of individual cells in the context of the movement the ripples within which they resided, by plotting cell reversals on the space-time DIC image of the corresponding ripples (see the gray-scale background of Figure 5-3 E). To make this image, a 2-D gray scale DIC images were rotated so that all the ripples travels parallel to their X-axis and then the ripples as observed by DIC imaging were averaged along the Y-axis (see 2.3.4). This space-time image illustrates the occurrence/timing and location of the wave crests (see the dark gray ridges in Figure 5-3 E). The colored lines are the trajectories of 6 GFP-labelled cells. Dots represent the points of cell reversals (the cells travel forward in-time, but can actually move forward or backward with reference to the x-axis [space]). By examining the overlapping cell trajectories of the cells within the context of the space-time DIC image of the ripples, the tracked cells travel with the high-density crests and reverse when and where two crests collide. Statistical analysis of position and time of cell reversals show that 75.0% (±2.6%) of all tracked cells reversal in the prey area occur during wave crests collisions. Similar

results were obtained from the ABM simulations (Figure 5-5). Interestingly, sometimes cells move through a counter-propagating wave crest without reversing and subsequently reverse with next crest. This pattern explains the small peak at about 12 min (about twice the average reversal time) in Figure 5-3 D. This second peak is even more obvious if the average distance travelled per reversal is plotted (Figure 5-6).

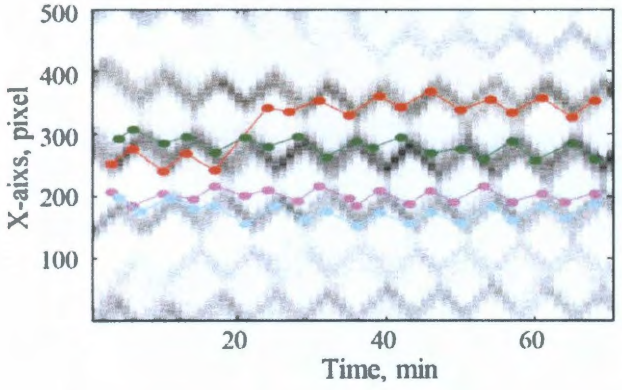


Figure 5-5 Space-time distribution of cell reversals in waves of ABM simulations. As in the experimental version of this figure (Figure 2E), the reversals of the agents usually (92% ±1.2%) occur during wave crest collisions, as all of the cells participate in rippling. For comparison, about 17% (±2.3%) of the reversals to occur in wave crests in a control population not on top of prey that does not sense side-to-side signaling.

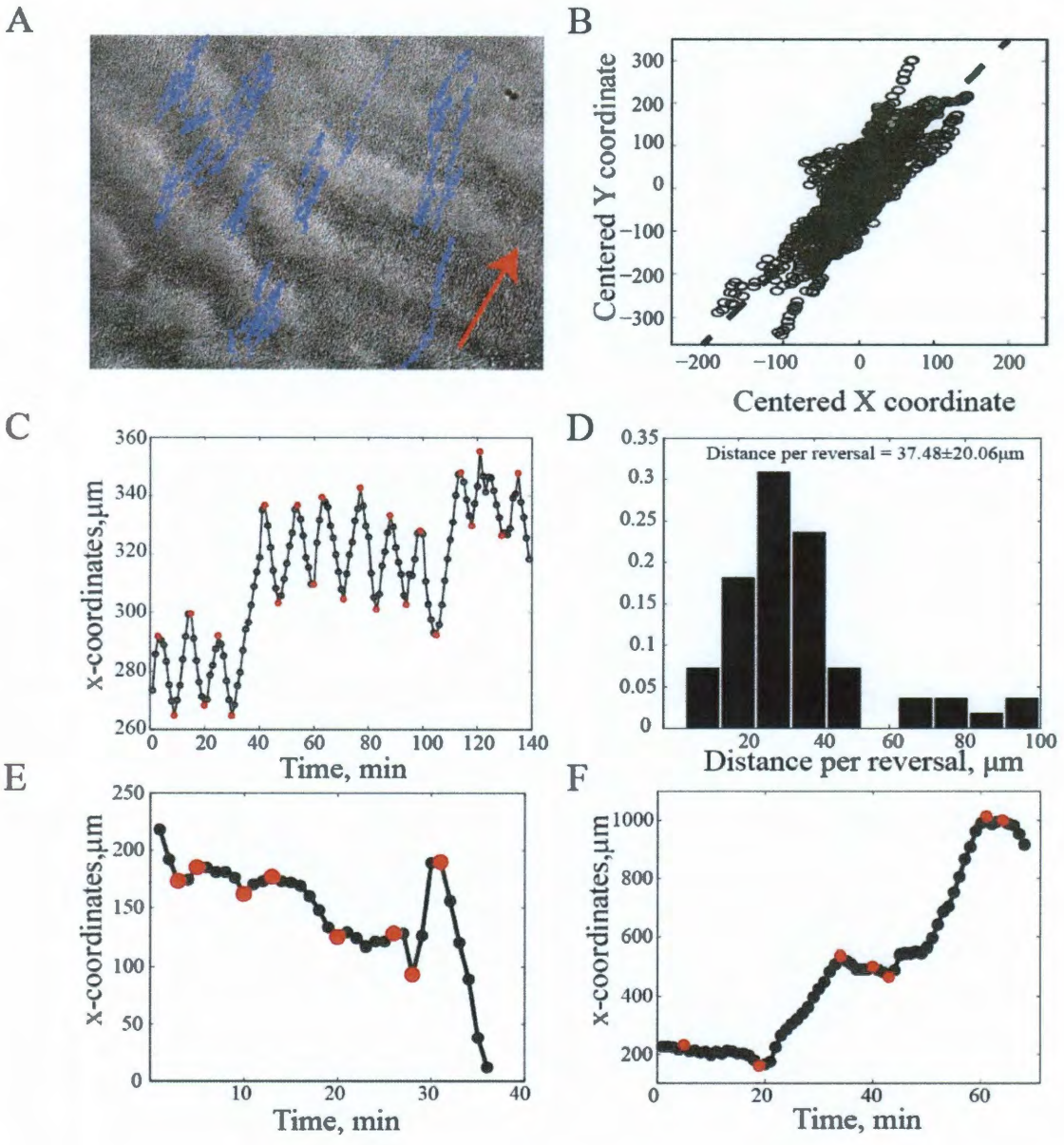


Figure 5-6 Results of experimental data analysis. (A) The background image is acquired from a DIC microscope which shows the ripple pattern. Individual cell trajectories of 11 cells are drawn as blue circles on the image. Both the background image and the cell trajectories are from the same movie. Cells mostly move along one direction, which is the same as the wave direction. The red arrow shows the direction of waves which is computed from principal component analysis (PCA). (C) A trajectory of a typical cell travel with ripple wave crest. The red dots denote where cellular reversals occur. (D) Distribution of distances that cells travel between reversals. (E) A trajectory of

a typical cell that is on prey but does not travel with wave crest. (F) A trajectory of a typical cell that is off prey.

5.2.4. Rippling facilitates cellular expansion into prey area

Even though significant attention has been devoted to the mechanisms of rippling, very little is known concerning the role of the phenomenon and its potential physiological benefits. It is previously hypothesized that developmental ripples may play a role in more uniform placement of fruiting bodies on the plate. However, recent observations indicate that rippling during the development may be coincidental and that rippling is mainly associated with the predatory behavior of *M. xanthus* [15, 17]. Interestingly, experimental results from Berelman et al. [15] indicate that when ripples form in the presence of prey, the expansion rate of the *M. xanthus* colony increases. Therefore this result is explored in the context of the predictions of the new ABM and experiments.

To test the effect of rippling interactions on the expansion rate with the new ABM, a simulation is conducted in which the cells that were predominantly aligned along their X-axis were placed into the central initial area from which they can expand to the right or to the left (Figure 5-7 top). To the right of the initial area a "prey area" was defined in which the cells effectively signal to one-another with side-by-side signaling and as a result form the ripples. To the left of the initial area the cells were ineffective in their side-by-side signaling. To quantify the cell expansion rate in simulations, I followed Wu et al. [29, 30], who used the observation of the constant density at the leading edge of the colony to conclude that the cell flux is linearly related to the steady state expansion rate. Therefore,

comparison of cell flux in both directions is chosen to assess the effect of rippling on the expansion rate. The simulation results (Figure 5-7) reveal a linear increase in the number of cells with time; the slope of this curve is a flux that is proportional to the cell expansion rate. The result of the least-square linear fit of the data show a 2.5-fold increase in the slope in the rippling region. Therefore, the ABM, which includes ripple-inducing interactions that increase the expansion rate, is in agreement with the observations of Berleman et al [15].

To test this prediction experimentally, the expansion rate of *M. xanthus* cells that were arranged on top of prey and not on top of prey was determined (Figure 5-7 bottom). The leading edges of the colony were measured over time and the expansion rates were calculated (see 2.3.6). The results reveal (Figure 5-7) that there is an increased expansion rate over prey of about 1.6-fold. This value is in qualitative agreement with the ABM prediction. The smaller increase of rippling expansion rate over prey measured experimentally is consistent with the hypothesis that not all the cells in the community participate in rippling behavior.

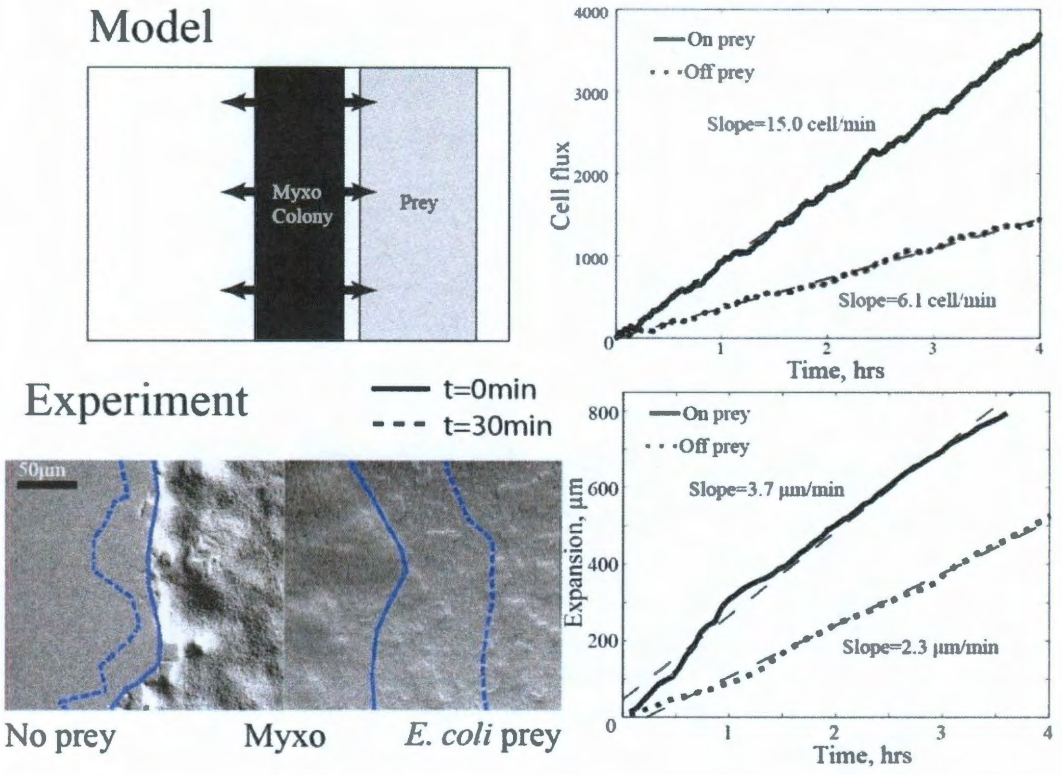


Figure 5-7 Ripples cause faster expansion of cells into the prey region. (A) ABM simulation was initialized according to this scheme. Agents that represent *M. xanthus* cells are placed in the black area. During the ABM simulations, the agents can move in both directions as the arrow indicates. On the right a gray area is defined to represent the prey area where agents can signal to one another and form ripples. (B) In the ABM simulations, cell flux is used to measure the expansion. The higher the cell flux, the faster the expansion rate. The higher cell flux in simulations on prey shows that *M. xanthus* colonies have a higher expansion rate on prey. (C) ImageJ software is used to track the edge of a *M. xanthus* colony and then calculated the rate of the edge movement as the expansion rate. The solid line represents the edge of *M. xanthus* colony in this image and the dash line indicates its location 30 min later. (D) The experimental observations of the expansion is plotted over time to show that the expansion rate over prey is about 1.6-fold larger than off the prey.

The rippling-dependent increased expansion rate is understandable in terms of the basic model ingredients. As the predatory cells expand over the prey from one side, there is a gradient of *M. xanthus* density, so each cell is less likely to encounter

a reversal-induced signaling event as they travel forward and more likely to encounter it as they move backward. This increases the bias in their motility and as a result they spread over the signaling region faster. Such behavior has clear physiological advantages in predation, as cells are able to relatively quickly spread over prey before their potential competitors.

5.2.5. Rippling keeps cells in the prey area for longer

Are there physiological benefits to rippling once *M. xanthus* cells completely and uniformly cover their prey? The simulation predicts that in this case, emerging traveling waves will decrease randomness in cellular motility and as a result the *M. xanthus* cells will be less likely to leave the prey regions. To quantitatively examine this prediction, a mean square displacement (MSD) of individual cells moving in the region is computed:

$$MSD = \left\langle (x_i - \langle x_i \rangle)^2 \right\rangle \qquad \text{Eq. (18)}$$

where angled brackets denote averaging. The results for cells in the regions with and without rippling interaction (Figure 5-8) are compared. As expected, the mean square displacement increased linearly with time as would be expected for random diffusion-like drift with no bias. Notably the slope of this curve (effective diffusion coefficient) is smaller for cells that are rippling (Figure 5-8 A, dotted-dash line); the effective diffusion coefficient for non-rippling is about 2.0-fold greater than for rippling cells. To ensure a well-controlled comparison between rippling and non-rippling cells, identical values are used for the average speeds, reversal periods and

all parameters regarding rippling noise. The results were further enhanced if the native (rather than average) reversal period is controlled (Figure 5-8 A, dotted-line). In this case, cells outside of the rippling region would reverse less frequently and therefore drift about 2.5 folds more than rippling cells

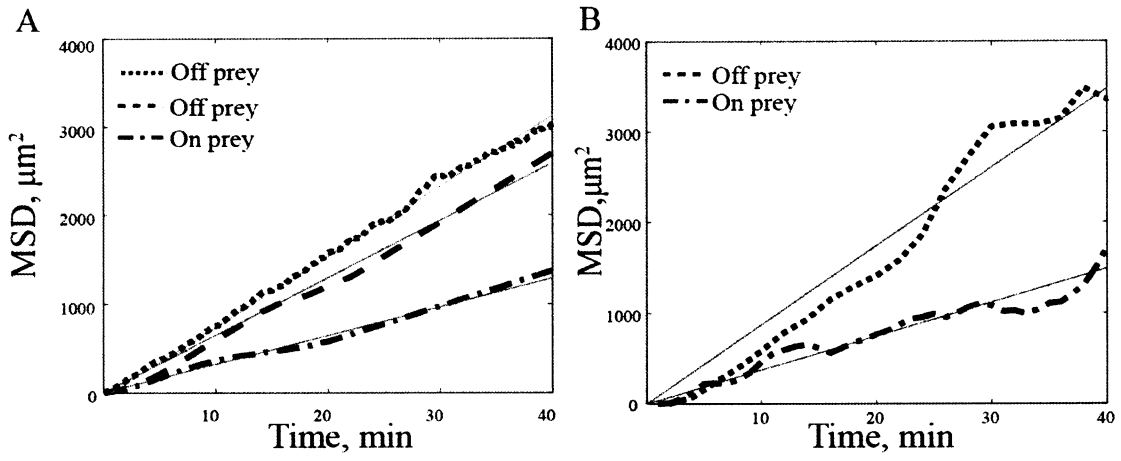


Figure 5-8 Mean square displacement (MSD) is used to characterize the random drift in cell movement. MSDs of *M. xanthus* cells on and off prey in both (A) ABM simulations and (B) experimental fluorescent images of cells on agar are compared. As a control, cells were placed on agar without prey, allowing them to move without side-to-side signaling, allowing them to mimic the behavior of cells off prey. For the experiments, ~100 cells were tracked both on and off prey and Eq. (3) was used to calculate the MSD by averaging the results for all cells. (A) shows the ABM simulation MSDs of cells on prey (dotted-dash line with a slope of 32.2 μm2/min), cells off prey with the same average reversal period as the cells on prey (dash line with a slope of 63.2 μm2/min), and cells off prey but with the same reversal period and velocity as cells off prey in experiments (dotted line with a slope of 81.5 μm2/min). (B) shows the experimental MSDs of cells on prey (dotted-dash line with a slope of 38.5 μm2/min), and cells off prey (dotted line with a slope of 85.5 μm2/min). The gray solid lines in both panels are no-intercept linear fit of the MSDs.

To test this prediction experimentally, GFP-labelled cells are tracked both in ripples (on top of prey) and outside the rippling region (not on top of prey). To

minimize the differences, the cells tracked at the same time and on the same plate, and Eq. (18) was used to compute MSD for each representative cell. The results (Figure 5-8 B) indicate that the observations are quantitatively similar to those predicted by the ABM (Figure 5-8 B), in that the rippling cells had lower drift and the non-rippling effective diffusion coefficient was about 2.2-fold larger than that for the rippling cells.

A decreased MSD is anticipated for rippling cells because the ripples introduce synchronized cell movement, so that the collective motility is less noisy than the individual cell motility. The rippling *M. xanthus* cells spend most of their time traveling back and forth within wave crests. They occasionally appear to skip a reversal, which allows them to join the next crest. Such synchronization also has a physiological advantage as a decrease in random drift makes it less likely that cells will move away from the prey accidentally.

5.3. Discussion

M. xanthus cells exhibit predatory rippling, which relies on the cells' ability to self-organize into periodic bands of traveling waves in response to the presence of prey bacteria. The mechanism and physiological role of predatory rippling has been defined here through the development of a new mathematical model that is based on and confirmed by direct observations. This new model is robust as it incorporates the recent concept of Mauriello et al. [59], that cellular reversals are triggered by a side-by-side signaling mechanism, with the formerly accepted ideas

that rippling results from the synchronized reversal of groups of cells moving in opposite directions[6-8, 36]. This new model reveals that three ingredients are sufficient to generate robust rippling behavior. The relationship between the wavelength, reversal time, and cell velocity could be formalized into a simple equation that describes rippling. The results confirm the observation of Berleman et al. [9, 10] that rippling behavior is a consequence of the contact with prey, which indirectly induce ripples by stimulating cell-to-cell signaling Finally, the results suggest that the physiological role of rippling behavior during *M. xanthus* predation is to increase the rate of spreading over prey cells and thereafter to allow cells to remain in contact with the prey for longer.

5.3.1. Mechanism of predatory rippling

Detailed characterization of predatory rippling by Berleman et al [9] has shown rippling to occur during predation on the variety of microorganisms and to be induced by the presence of macromolecular substances. These observations, which link rippling with the presence of macromolecules, led Berelman et al. [9] to suggest that ripples originate solely as a interactions of individual cells with macromolecules and to neglect any self-organizing intercellular interactions. In contrast to this concept, it is proposed here that ripples result from the self-organization of cells into traveling wave patterns, which result from intercellular signaling that is stimulated or facilitated by the presence of macromolecules. Indeed, in experimental arrangement the macromolecules are likely to be distributed uniformly and their concentration is expected to vary very little during the typical

wave period (~10 min). Moreover, even if macromolecules induce periodicity of *M. xanthus* cell motility as suggested by Berleman et al. this would not be sufficient to induce ripples because their formation requires temporal and spatial synchronization of cellular behavior unattainable without cell-to-cell signaling.

Previous mathematical models of rippling suggested that such self-organization occurs when counter-moving cells interact to induce reversals [7, 36]. For developmental rippling this interaction was proposed to rely on the starvation-induced C-signal, a 17-kDa form of the CsgA protein, presumably localized to the cell poles and resulting in signaling events between two cells during their pole-to-pole collisions. However, recent observation of Mauriello et al [59] suggested a novel more robust mechanism for intercellular signaling. When counter-moving cells come into side-to-side contact, clusters of chemotactic FrzCD receptors within the cells show transient alignments and thereafter one of the cells reverses. The new mathematical model shows that this side-to-side signaling is sufficient to induce rippling self-organization in a locally aligned cell population, assuming that there is a minimal time duration of a refractory period, during which cells can not reverse again regardless of signaling. The existence of the refractory period has also been assumed in the previous model [7, 36] and this assumption is very plausible as reversal is anticipated to require a significant reorganization of the cell-motility machinery[43, 149, 150]. The existence of a refractory period also naturally follows from dynamic properties of a negative-feedback oscillator (Frz-ilator), which was hypothesized to regulate cell reversals. Altogether, the modeling results suggest that the self-organization of cells into ripples during predation can be explained by

the increased efficiency or higher probability of side-to-side signaling induced by the presence of prey macromolecules.

Are there differences in the mechanisms of the predatory and developmental ripples? Mathematically, the new model of rippling with side-to-side signaling is very similar to the older pole-to-pole signaling model and therefore is expected to be in good agreement with the experimental patterns that were previously observed [6, 8]. Moreover, the testing of the side-to-side signaling model has shown it to be much more robust, in that it can withstand realistic levels of variability in cell speed and reversal times (Figure 5-2). Furthermore, the experiments of Berleman et al [9, 10] showed evidence that developmental rippling occurs as a side effect of cell lysis during the aggregation. The new side-to-side signaling model of rippling should be applicable to describe rippling in both conditions. It remains to be seen however, how the presence of C-signaling affects the FrzCD cluster alignment and whether C-signaling plays any role in rippling during predation.

5.3.2. Quantitative and qualitative agreement between the new model and experimental observations

Although the main hypothesis of this model, which is that rippling behavior is initiated by the side-to-side signaling by the presence of prey cells, cannot be directly tested at this time, due to lack of an understanding of the underlying biochemical mechanisms, one can nevertheless compare the model predictions with experimental data collected by this work and others and thereby rigorously test the model. One obvious consequence of the assumption would be more frequent

reversals of *M. xanthus* cells when prey is present. This agrees with experimental observation of Berleman et al. [10] and the new data. Moreover the resulting self-organization of cells into ripples provides various ways to quantitatively and qualitatively compare *in silico* generated ripples with experimental observations. If the presence of prey stimulates this signaling, then the rippling would only be observed at the regions where signaling is sufficiently probable, i.e. only in the regions covered with prey in agreement with experimental observations in my work and those of Berleman et al. [9, 10]. Indeed, simulations show that the signaling probability can serve as a bifurcation parameter that leads to a transition between the homogeneous cell distribution and the formation of ripples (Appendix A).

The timescale of rippling self-organization can be defined as the time it takes to generate waves starting from an initially homogeneous cell population into to well-focused wave patterns. The model predicts that time to be of the order of 3 h, which is remarkably consistent with experimental observations (Figure 5-1). Qualitative comparison of the time-lapse dynamics also shows a very good agreement. Interestingly, the time-scale of wave origination in the experiments of Berleman et al is significantly longer ~12 h. Although it is hard to pin-point the source of this discrepancy, the new model indicates that the cell density and the amount of noise in cell orientation can significantly affect wave-synchronization time.

Measuring rippling wavelengths and correlating them to the parameters of individual cell motility provides another opportunity to test the mathematical

model. This analysis reveals a slightly modified relationship between the wavelength, wave-crest width, individual cell speed and reversal time than has been previously established [7, 36] (Eq. (17)). This relationship is confirmed by simulations and is in excellent agreement with the experimental measurements of wavelength (Figure 5-3A and B). The wavelength prediction is also compatible with previously reported measurements [10] and with observations of Sliusarenko et al [9] that show that cells moving in opposite directions tend to interpenetrate one cell length before reversal is triggered. Figure 5-4shows the sequence of events that occur during two-crest collisions. This cartoon model indicates that once the cells in front of each crest reverse they would signal to the cells that were following them, resulting in chain-reaction of signaling and reversal events. This cartoon also illustrates the importance of the refractory period – once the cell in front of the crest reverses, it is essential for it to keep signaling to other cells to reverse without reversing itself. Further confirmation of the model comes from tracking the cell reversals in time and space and comparing these with the locations of wave-crest collisions. The experimental results (Figure 5-3 E) indicate that most reversals occur when two wave crests collide, just as the model predicts (Figure 5-6 F).

5.3.3. Physiological role of rippling

The previous model of developmental ripples suggested [31] that periodic travelling waves can ensure more regular distribution of fruiting-body aggregates at the colony edge, as seen in the submerged culture setup of Welch et al[8]. However, the physiological implications of this observation are unclear as ordering of

aggregate distribution can occur even without rippling [23]. And if rippling is predominantly triggered during the predation what is its role in this behavioral phenomenon? Berleman et al. [151] proposed two hypotheses. The first, termed the "grinder model" speculates that the movement of the waves of *M. xanthus* cells during rippling causes a physical disruption of the prey colony. The second, termed the "population control model" suggests that waves maximize the prey-predator contact area and push the excess of predator cells to the edges of rippling area. Neither of these hypotheses is likely to be correct, based on the biophysics of their environment in which there is a very-low Reynolds number and hydrodynamics does not allow mixing or transport by temporary periodic perturbations (ref). Nevertheless, the mathematical model suggests several alterative physiological benefits of rippling for predation. These predictions are consistent with the experimental observations reported here and previously.

First, during the expansion over prey, the presence of rippling-inducing side-to-side signaling significantly facilitates the rate of *M. xanthus* cell spreading (Figure 3). As a result cells will cover their prey faster, which has obvious physiological benefits and probably an advantage in the competitive soil environment. This observation is consistent with the results reported earlier by Berleman et al [10] and with experimental results in my work. Notably, this result does not require ripples per se but only reversal inducing signaling. However, as the model shows that signaling is key for rippling self-organization and the other model ingredients can be easily justified by what is known about the biophysics of *M. xanthus* motility. The increase in spreading also requires a gradient in the density of *M. xanthus* cells,

which is always present in the leading edge of the swarm. On the other hand, the observation does not require a gradient in the density of prey as the alternative chemotaxis-based explanation would predict.

Second, we've predicted that cells that ripple in the absence of a cell-density gradient (i.e. when they are behind the leading edge of the swarm or once the prey is fully covered), would engage in less noisy and more periodic motion and as a result will have less of a random drift (Figure 4). This effect would help cells to stay in the prey area for longer and prevent them from accidentally moving away. The prediction is confirmed in the cell-tracking assays and is consistent with previous unquantified observations of Berleman et al [10]. Notably, this effect requires waves to form, as it is collective interaction of cells in the waves that leads to their synchronization. The effect is in many ways analogous to the well-known mathematical phenomena in which a collection of coupled noisy oscillators is less noisy than each oscillator on its own [152].

In addition, it is possible that the formation of the ripples would also increase cell alignment as steric interactions in the denser crests will be more efficient. This prediction agrees with observation of Berleman et al. [10] and my own observations. On the other hand it is worth noting that the causal relationship between rippling and alignment is not obvious, as ripples also require cell alignment. Therefore, it is likely that there is a positive self-reinforcing feedback loop between the formation of ripples and cell alignment: as cells align, ripples become more pronounced and their crests become more dense leading to further

cell alignment. Although the physiological benefit of better alignment is not obvious it may further enhance the rate of spreading contributing to the effect discuss above.

Chapter 6

General Conclusions

Self-organization in microbial colonies such as *M. xanthus* fruiting body development and ripple formation are complex emergent behaviors resulting from coordinated interactions of thousands of cells controlled on multiple levels. Mechanistic understanding of such coordination is not evident from qualitative observations and needs to be supplemented with quantitative approaches. With combination of agent-based modeling, time-lapse fluorescence microscopy experiments, image quantification and high-throughput statistical analysis presented in this thesis new insights have been gained on the mechanisms of *M. xanthus* self-organization into ripples during predation and aggregation into fruiting bodies under starvation condition.

The combination of the approaches is undoubtedly synergistic. The agent-based model demonstrates that a simple set of ingredients based on the experimental observation is sufficient to produce rippling patterns which are

identical to experimental observed patterns. The agent-based model for aggregation creates quantitative match of patterns of steady state between simulations and experiments. However, demonstrated mismatch for the transient patterns indicates that the traffic jam mechanism alone is not sufficient to accurately reproduce the later stages of aggregation. The agent-based model not only generates means to quickly test the effects of hypothesized intercellular interactions on collective motility, it provides the straight forward representation of the intercellular dynamics, social interaction among groups of microorganisms as well.

The subsequent experiments test the variety of model predictions and allow refining the model to achieve quantitative agreement with the new data. Statistical image processing is the key for characterization of experimental results. The statistical image processing can quantitatively compare results from both simulations and experiments and provide parameter values that are used in agent-based models. The statistical image processing developed can be further extended and adapted to study self-organization in other bacteria or unicellular eukaryotes. This is essential in order to quantitatively compare experimental patterns with their *in silico* counterparts as well as to statistically assess biological hypotheses on self-organization dynamics.

The combination of approaches is essential to further the understanding self-organization in more complex systems such as development of multicellular organisms.

References

1. Reichenbach, H., *The ecology of the myxobacteria.* Environ Microbiol, 1999. **1**(1): p. 15-21.
2. Hillesland, H., et al., *Identification of aerobic gut bacteria from the kala azar vector, Phlebotomus argentipes: a platform for potential paratransgenic manipulation of sand flies.* Am J Trop Med Hyg, 2008. **79**(6): p. 881-6.
3. Konovalova, A., T. Petters, and L. Sogaard-Andersen, *Extracellular biology of Myxococcus xanthus.* FEMS Microbiol Rev. **34**(2): p. 89-106.
4. Henrichsen, J., *Bacterial surface translocation: a survey and a classification.* Bacteriol Rev, 1972. **36**(4): p. 478-503.
5. Zusman, D.R., et al., *Chemosensory pathways, motility and development in Myxococcus xanthus.* Nat Rev Microbiol, 2007. **5**(11): p. 862-72.
6. Sager, B. and D. Kaiser, *Intercellular C-signaling and the traveling waves of Myxococcus.* Genes Dev, 1994. **8**(23): p. 2793-804.
7. Igoshin, O.A., et al., *Pattern formation and traveling waves in myxobacteria: theory and modeling.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 14913-8.
8. Welch, R. and D. Kaiser, *Cell behavior in traveling wave patterns of myxobacteria.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 14907-12.
9. Berleman, J.E., et al., *Rippling is a predatory behavior in Myxococcus xanthus.* J Bacteriol, 2006. **188**(16): p. 5888-95.
10. Berleman, J.E., et al., *Predataxis behavior in Myxococcus xanthus.* Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17127-32.
11. Berleman, J.E. and J.R. Kirby, *Multicellular development in Myxococcus xanthus is stimulated by predator-prey interactions.* J Bacteriol, 2007. **189**(15): p. 5675-82.
12. Dworkin, M., *Nutritional requirements for vegetative growth of Myxococcus xanthus.* J Bacteriol, 1962. **84**: p. 250-7.
13. Dworkin, M. and H. Voelz, *The formation and germination of microcysts in Myxococcus xanthus.* J Gen Microbiol, 1962. **28**: p. 81-5.
14. Kuner, J.M. and D. Kaiser, *Fruiting body morphogenesis in submerged cultures of Myxococcus xanthus.* J Bacteriol, 1982. **151**(1): p. 458-61.
15. Varon, M., S. Cohen, and E. Rosenberg, *Autocides produced by Myxococcus xanthus.* J Bacteriol, 1984. **160**(3): p. 1146-50.
16. Velicer, G.J., L. Kroos, and R.E. Lenski, *Developmental cheating in the social bacterium Myxococcus xanthus.* Nature, 2000. **404**(6778): p. 598-601.
17. Velicer, G.J. and Y.T. Yu, *Evolution of novel cooperative swarming in the bacterium Myxococcus xanthus.* Nature, 2003. **425**(6953): p. 75-8.
18. Fiegna, F., et al., *Evolution of an obligate social cheater to a superior cooperator.* Nature, 2006. **441**(7091): p. 310-4.
19. Pelling, A.E., et al., *Nanoscale visualization and characterization of Myxococcus xanthus cells with atomic force microscopy.* Proc Natl Acad Sci U S A, 2005. **102**(18): p. 6484-9.

20. Kaiser, D. and H. Warrick, *Myxococcus xanthus swarms are driven by growth and regulated by a pacemaker.* J Bacteriol, 2011. **193**(21): p. 5898-904.
21. Shimkets, L. and T.W. Seale, *Fruiting-body formation and myxospore differentiation and germination in Mxyococcus xanthus viewed by scanning electron microscopy.* J Bacteriol, 1975. **121**(2): p. 711-20.
22. Xie, C., et al., *Statistical image analysis reveals features affecting fates of Myxococcus xanthus developmental aggregates.* Proc Natl Acad Sci U S A. **108**(14): p. 5915-20.
23. Zhang, H., et al., *Quantifying aggregation dynamics during Myxococcus xanthus development.* J Bacteriol.
24. Kaiser, D. and C. Crosby. *Cell movement and its coordination in swarms of myxococcus xanthus.* 1983.
25. O'Connor, K.A. and D.R. Zusman, *Patterns of cellular interactions during fruiting-body formation in Myxococcus xanthus.* J Bacteriol, 1989. **171**(11): p. 6013-24.
26. Otani, M., M. Inouye, and S. Inouye, *Germination of myxospores from the fruiting bodies of Myxococcus xanthus.* J Bacteriol, 1995. **177**(15): p. 4261-5.
27. Othmer, H.G. and A. Stevens, *Aggregation, blowup, and collapse: The ABC's of taxis in reinforced random walks.* Siam Journal on Applied Mathematics, 1997. **57**(4): p. 1044-1081.
28. Jelsbak, L. and L. Sogaard-Andersen, *Pattern formation: fruiting body morphogenesis in Myxococcus xanthus.* Curr Opin Microbiol, 2000. **3**(6): p. 637-42.
29. Camazine, S., et al., *Self-Organization in Biological Systems.* 2001, Princeton. NJ: Princeton University Press,.
30. Igoshin, O.A., D. Kaiser, and G. Oster, *Breaking symmetry in myxobacteria.* Curr Biol, 2004. **14**(12): p. R459-62.
31. Igoshin, O.A., J. Neu, and G. Oster, *Developmental waves in myxobacteria: A distinctive pattern formation mechanism.* Phys Rev E Stat Nonlin Soft Matter Phys, 2004. **70**(4 Pt 1): p. 041911.
32. Igoshin, O.A., et al., *Waves and aggregation patterns in myxobacteria.* Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4256-61.
33. Kaiser, D. and R. Welch, *Dynamics of fruiting body morphogenesis.* J Bacteriol, 2004. **186**(4): p. 919-27.
34. Sozinova, O., et al., *A three-dimensional model of myxobacterial aggregation by contact-mediated interactions.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11308-12.
35. Stevens, A. and L. Sogaard-Andersen, *Making waves: pattern formation by a cell-surface-associated signal.* Trends Microbiol, 2005. **13**(6): p. 249-52.
36. Sliusarenko, O., et al., *Accordion waves in Myxococcus xanthus.* Proc Natl Acad Sci U S A, 2006. **103**(5): p. 1534-9.
37. Curtis, P.D., et al., *Spatial organization of Myxococcus xanthus during fruiting body formation.* J Bacteriol, 2007. **189**(24): p. 9126-30.

38. Kaiser, D., *Bacterial swarming: a re-examination of cell-movement patterns.* Curr Biol, 2007. **17**(14): p. R561-70.
39. Sliusarenko, O., D.R. Zusman, and G. Oster, *Aggregation during fruiting body formation in Myxococcus xanthus is driven by reducing cell movement.* J Bacteriol, 2007. **189**(2): p. 611-9.
40. Wu, Y., et al., *Periodic reversal of direction allows Myxobacteria to swarm.* Proc Natl Acad Sci U S A, 2009. **106**(4): p. 1222-7.
41. Holmes, A.B., S. Kalvala, and D.E. Whitworth, *Spatial simulations of myxobacterial development.* PLoS Comput Biol, 2010. **6**(2): p. e1000686.
42. Mauriello, E.M., *Cell polarity/motility in bacteria: closer to eukaryotes than expected?* Embo J. **29**(14): p. 2258-9.
43. Nan, B., et al., *Myxobacteria gliding motility requires cytoskeleton rotation powered by proton motive force.* Proc Natl Acad Sci U S A. **108**(6): p. 2498-503.
44. Hodgkin, J. and D. Kaiser, *Cell-to-cell stimulation of movement in nonmotile mutants of Myxococcus.* Proc Natl Acad Sci U S A, 1977. **74**(7): p. 2938-42.
45. Kaiser, D., *Social gliding is correlated with the presence of pili in Myxococcus xanthus.* Proc Natl Acad Sci U S A, 1979. **76**(11): p. 5952-6.
46. Zusman, D.R., *"Frizzy" mutants: a new class of aggregation-defective developmental mutants of Myxococcus xanthus.* J Bacteriol, 1982. **150**(3): p. 1430-7.
47. LaRossa, R., et al., *Developmental cell interactions of Myxococcus xanthus: analysis of mutants.* J Bacteriol, 1983. **153**(3): p. 1394-404.
48. Kim, S.K. and D. Kaiser, *Cell alignment required in differentiation of Myxococcus xanthus.* Science, 1990. **249**(4971): p. 926-8.
49. Li, S., B.U. Lee, and L.J. Shimkets, *csgA expression entrains Myxococcus xanthus development.* Genes Dev, 1992. **6**(3): p. 401-10.
50. Munoz-Dorado, J. and J.M. Arias, *The social behavior of myxobacteria.* Microbiologia, 1995. **11**(4): p. 429-38.
51. Spormann, A.M. and A.D. Kaiser, *Gliding movements in Myxococcus xanthus.* J Bacteriol, 1995. **177**(20): p. 5846-52.
52. Spormann, A.M. and D. Kaiser, *Gliding mutants of Myxococcus xanthus with high reversal frequencies and small displacements.* J Bacteriol, 1999. **181**(8): p. 2593-601.
53. Kaiser, D., *Bacterial motility: how do pili pull?* Curr Biol, 2000. **10**(21): p. R777-80.
54. Kruse, T., et al., *C-signal: a cell surface-associated morphogen that induces and co-ordinates multicellular fruiting body morphogenesis and sporulation in Myxococcus xanthus.* Mol Microbiol, 2001. **40**(1): p. 156-68.
55. Wolgemuth, C., et al., *How myxobacteria glide.* Curr Biol, 2002. **12**(5): p. 369-77.
56. Igoshin, O.A., et al., *A biochemical oscillator explains several aspects of Myxococcus xanthus behavior during development.* Proc Natl Acad Sci U S A, 2004. **101**(44): p. 15760-5.

57. Mignot, T., *The elusive engine in Myxococcus xanthus gliding motility.* Cell Mol Life Sci, 2007. **64**(21): p. 2733-45.
58. Mignot, T., et al., *Evidence that focal adhesion complexes power bacterial gliding motility.* Science, 2007. **315**(5813): p. 853-6.
59. Mauriello, E.M., et al., *Localization of a bacterial cytoplasmic receptor is dynamic and changes with cell-cell contacts.* Proc Natl Acad Sci U S A, 2009. **106**(12): p. 4852-7.
60. Couzin, I.D., et al., *Collective memory and spatial sorting in animal groups.* J Theor Biol, 2002. **218**(1): p. 1-11.
61. Nadell, C.D., K.R. Foster, and J.B. Xavier, *Emergence of spatial structure in cell groups and the evolution of cooperation.* PLoS Comput Biol, 2010. **6**(3): p. e1000716.
62. Anderson, A.R. and B.N. Vasiev, *An individual based model of rippling movement in a myxobacteria population.* J Theor Biol, 2005. **234**(3): p. 341-9.
63. McBride, M.J., *Bacterial gliding motility: multiple mechanisms for cell movement over surfaces.* Annu Rev Microbiol, 2001. **55**: p. 49-75.
64. Hodgkin, J. and D. Kaiser, *Genetics of gliding motility in <i>Myxococcus xanthus</i> (Myxobacterales): Two gene systems control movement.* Molecular and General Genetics MGG, 1979. **171**(2): p. 177-191.
65. Hodgkin, J. and D. Kaiser, *Genetics of gliding motility in <i>Myxococcus xanthus</i> (Myxobacterales): Genes controlling movement of single cells.* Molecular and General Genetics MGG, 1979. **171**(2): p. 167-176.
66. Kaiser, D. and C. Crosby, *Cell-Movement and Its Coordination in Swarms of Myxococcus-Xanthus.* Cell Motility and the Cytoskeleton, 1983. **3**(3): p. 227-245.
67. Kaiser, D. [cited 2011; Available from: http://cmgm.stanford.edu/devbio/kaiserlab/about_myxo/about_myxococcus.html.
68. Blackhart, B.D. and D.R. Zusman, *"Frizzy" genes of Myxococcus xanthus are involved in control of frequency of reversal of gliding motility.* Proc Natl Acad Sci U S A, 1985. **82**(24): p. 8767-70.
69. Jahn, E., *Beiträge zur botanischen protistologie.* 1924: Gebrüder Borntraeger.
70. Nan, B. and D.R. Zusman, *Uncovering the mystery of gliding motility in the myxobacteria.* Annu Rev Genet, 2011. **45**: p. 21-39.
71. Sun, H., Z. Yang, and W. Shi, *Effect of cellular filamentation on adventurous and social gliding motility of Myxococcus xanthus.* Proc Natl Acad Sci U S A, 1999. **96**(26): p. 15178-83.
72. Sliusarenko, O., D.R. Zusman, and G. Oster, *The motors powering A-motility in Myxococcus xanthus are distributed along the cell body.* J Bacteriol, 2007. **189**(21): p. 7920-1.
73. Bowden, M.G. and H.B. Kaplan, *The Myxococcus xanthus lipopolysaccharide O-antigen is required for social motility and multicellular development.* Mol Microbiol, 1998. **30**(2): p. 275-84.

74. Li, Y., et al., *Extracellular polysaccharides mediate pilus retraction during social motility of Myxococcus xanthus.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5443-8.
75. Shi, W. *Myxococcus xanthus.* [cited 2011; Available from: http://www.mimg.ucla.edu/faculty/shi/Myxo.htm.
76. Blackhart, B.D. and D.R. Zusman, *Analysis of the products of the Myxococcus xanthus frz genes.* J Bacteriol, 1986. **166**(2): p. 673-8.
77. Shi, W. and D.R. Zusman, *Methionine inhibits developmental aggregation of Myxococcus xanthus by blocking the biosynthesis of S-adenosyl methionine.* J Bacteriol, 1995. **177**(18): p. 5346-9.
78. Shimkets, L.J. and D. Kaiser, *Induction of coordinated movement of Myxococcus xanthus cells.* J Bacteriol, 1982. **152**(1): p. 451-61.
79. Pennisi, E., *How did cooperative behavior evolve.* Science, 2005. **309**(5731): p. 93-93.
80. Vogel, G., *How does a single somatic cell become a whole plant.* Science, 2005. **309**(5731): p. 86-86.
81. Tomlin, C.J. and J.D. Axelrod, *Biology by numbers: mathematical modelling in developmental biology.* Nat Rev Genet, 2007. **8**(5): p. 331-40.
82. Stoodley, P., et al., *Biofilms as complex differentiated communities.* Annu Rev Microbiol, 2002. **56**: p. 187-209.
83. Danese, P.N., L.A. Pratt, and R. Kolter, *Biofilm formation as a developmental process.* Methods Enzymol, 2001. **336**: p. 19-26.
84. O'Toole, G., H.B. Kaplan, and R. Kolter, *Biofilm formation as microbial development.* Annu Rev Microbiol, 2000. **54**: p. 49-79.
85. Ben-Jacob, E., I. Cohen, and H. Levine, *Cooperative self-organization of microorganisms.* Advances in Physics, 2000. **49**(4): p. 395-554.
86. Goldman, B.S., et al., *Evolution of sensory complexity recorded in a myxobacterial genome.* Proc Natl Acad Sci U S A, 2006. **103**(41): p. 15200-5.
87. Kroos, L., *Eukaryotic-like signaling and gene regulation in a prokaryote that undergoes multicellular development.* Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2681-2.
88. Kaiser, D., *Coupling cell movement to multicellular development in myxobacteria.* Nat Rev Microbiol, 2003. **1**(1): p. 45-54.
89. Kaiser, D., *Signaling in myxobacteria.* Annu Rev Microbiol, 2004. **58**: p. 75-98.
90. McVittie, A., F. Messik, and S.A. Zahler, *Developmental biology of Myxococcus.* J Bacteriol, 1962. **84**: p. 546-51.
91. Koch, A.L., *Unidirectional movement of flares of cells of Myxococcus xanthus.* Crit Rev Microbiol, 2006. **32**(2): p. 87-90.
92. Sager, B.M., *Propagation of traveling waves in excitable media.* Genes Dev, 1996. **10**(18): p. 2237-50.
93. Wireman, J., *Developmental induction of Myxococcus xanthus myxospores.* J Bacteriol, 1979. **140**(1): p. 147-53.

94. Fremgen, S.A., N.S. Burke, and P.L. Hartzell, *Effects of site-directed mutagenesis of mglA on motility and swarming of Myxococcus xanthus.* BMC Microbiol. **10**: p. 295.
95. Wu, Y., et al., *Social interactions in myxobacterial swarming.* PLoS Comput Biol, 2007. **3**(12): p. e253.
96. Ermentrout, G.B. and L. Edelstein-Keshet, *Cellular automata approaches to biological modeling.* Journal of theoretical Biology, 1993. **160**: p. 97-97.
97. Deutsch, A., S. Dormann, and P.K. Maini, *Cellular automaton modeling of biological pattern formation: characterization, applications, and analysis.* 2005: Springer.
98. Chopard, B., M. Droz, and C.U. Press, *Cellular automata modeling of physical systems*. 1998: Cambridge University Press Cambridge^ eMA MA.
99. Wolfram, S., *Theory and applications of cellular automata.* 1986.
100. Bonabeau, E., *Agent-based modeling: methods and techniques for simulating human systems.* Proc Natl Acad Sci U S A, 2002. **99 Suppl 3**: p. 7280-7.
101. Niazi, M. and A. Hussain, *Agent-based computing from multi-agent systems to agent-based models: a visual survey.* Scientometrics, 2011. **89**(2): p. 479-499.
102. Gjerdrum, J., N. Shah, and L.G. Papageorgiou, *A combined optimization and agent-based approach to supply chain modelling and performance assessment.* Production Planning & Control, 2001. **12**(1): p. 81-88.
103. Schenk, T.A., G. Loffler, and J. Rauh, *Agent-based simulation of consumer behavior in grocery shopping on a regional level.* Journal of Business research, 2007. **60**(8): p. 894-903.
104. Watkins, M., et al., *Using agent-based modeling to study construction labor productivity as an emergent property of individual and crew interactions.* Journal of Construction Engineering and Management, 2009. **135**: p. 657.
105. Erol, K., R. Levy, and J. Wentworth, *Application of agent technology to traffic simulation.* 2000.
106. Van den Brink, P.J., et al., *An individual-based approach to model spatial population dynamics of invertebrates in aquatic ecosystems after pesticide contamination.* Environmental Toxicology and Chemistry, 2007. **26**(10): p. 2226-2236.
107. Ratze, C., et al., *Simulation modelling of ecological hierarchies in constructive dynamical systems.* Ecological complexity, 2007. **4**(1-2): p. 13-25.
108. An, G., *Agent-based computer simulation and sirs: building a bridge between basic science and clinical trials.* Shock (Augusta, Ga.), 2001. **16**(4): p. 266.
109. Mansury, Y., et al., *Emerging patterns in tumor systems: simulating the dynamics of multicellular clusters with an agent-based spatial agglomeration model.* Journal of theoretical Biology, 2002. **219**(3): p. 343-370.
110. Walker, D.C., et al., *Agent-based computational modeling of wounded epithelial cell monolayers.* IEEE Trans Nanobioscience, 2004. **3**(3): p. 153-63.
111. Peirce, S.M., E.J. Van Gieson, and T.C. Skalak, *Multicellular simulation predicts microvascular patterning and in silico tissue assembly.* The FASEB journal, 2004. **18**(6): p. 731-733.

112. Pogson, M., et al., *Formal agent-based modelling of intracellular chemical interactions.* Biosystems, 2006. **85**(1): p. 37-45.
113. Auchincloss, A.H. and A.V.D. Roux, *A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health.* American journal of epidemiology, 2008. **168**(1): p. 1-8.
114. Folcik, V.A., G.C. An, and C.G. Orosz, *The Basic Immune Simulator: an agent-based model to study the interactions between innate and adaptive immunity.* Theor Biol Med Model, 2007. **4**: p. 39.
115. *Repast.* [cited 2011; Available from: http://repast.sourceforge.net/repast_3/index.html.
116. Java. *Java concurrency.* [cited 2011; Available from: http://docs.oracle.com/javase/tutorial/essential/concurrency/index.html.
117. RiceUniversity. [cited 2011; Available from: http://www.rcsg.rice.edu/home/.
118. Taylor, R.G. and R.D. Welch, *Recording multicellular behavior in Myxococcus xanthus biofilms using time-lapse microcinematography.* J Vis Exp, 2010(42).
119. D'Errico, J., *Surface Fitting using gridfit.* 2005, MATLAB Central File Exchange.
120. Felzenszwalb, P.F. and D.P. Huttenlocher, *Efficient belief propagation for early vision.* International Journal of Computer Vision, 2006. **70**(1): p. 41-54.
121. D'Errico, J. *Surface Fitting using gridfit.* 2005; Available from: http://www.mathworks.com/matlabcentral/fileexchange/8998.
122. Cover, T.M. and J.A. Thomas, *Elements of information theory*. 2nd ed. 2006, Hoboken, N.J.: Wiley-Interscience. xxiii, 748 p.
123. Burges, C.J.C., *A tutorial on Support Vector Machines for pattern recognition.* Data Mining and Knowledge Discovery, 1998. **2**(2): p. 121-167.
124. Nariya, H. and M. Inouye, *MazF, an mRNA interferase, mediates programmed cell death during multicellular Myxococcus development.* Cell, 2008. **132**(1): p. 55-66.
125. ImageJ. [cited 2010; Available from: http://rsbweb.nih.gov/ij/index.html.
126. Morlet, J., et al., *Wave propagation and sampling theory---Part II: Sampling theory and complex waves.* Geophysics, 1982. **47**(2): p. 222-236.
127. Mallat, S.G., *Multiresolution Approximations and Wavelet Orthonormal Bases of L2(R).* Transactions of the American Mathematical Society, 1989. **315**(1): p. 69-87.
128. Mallat, S.G., *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation.* IEEE Trans. Pattern Anal. Mach. Intell., 1989. **11**(7): p. 674-693.
129. Werner, R., et al., *Application of wavelet transformation to determine wavelengths and phase velocities of gravity waves observed by lidar measurements.* Journal of Atmospheric and Solar-Terrestrial Physics, 2007. **69**(17-18): p. 2249-2256.
130. Daubechies, I., *The wavelet transform, time-frequency localization and signal analysis.* Information Theory, IEEE Transactions on, 1990. **36**(5): p. 961-1005.

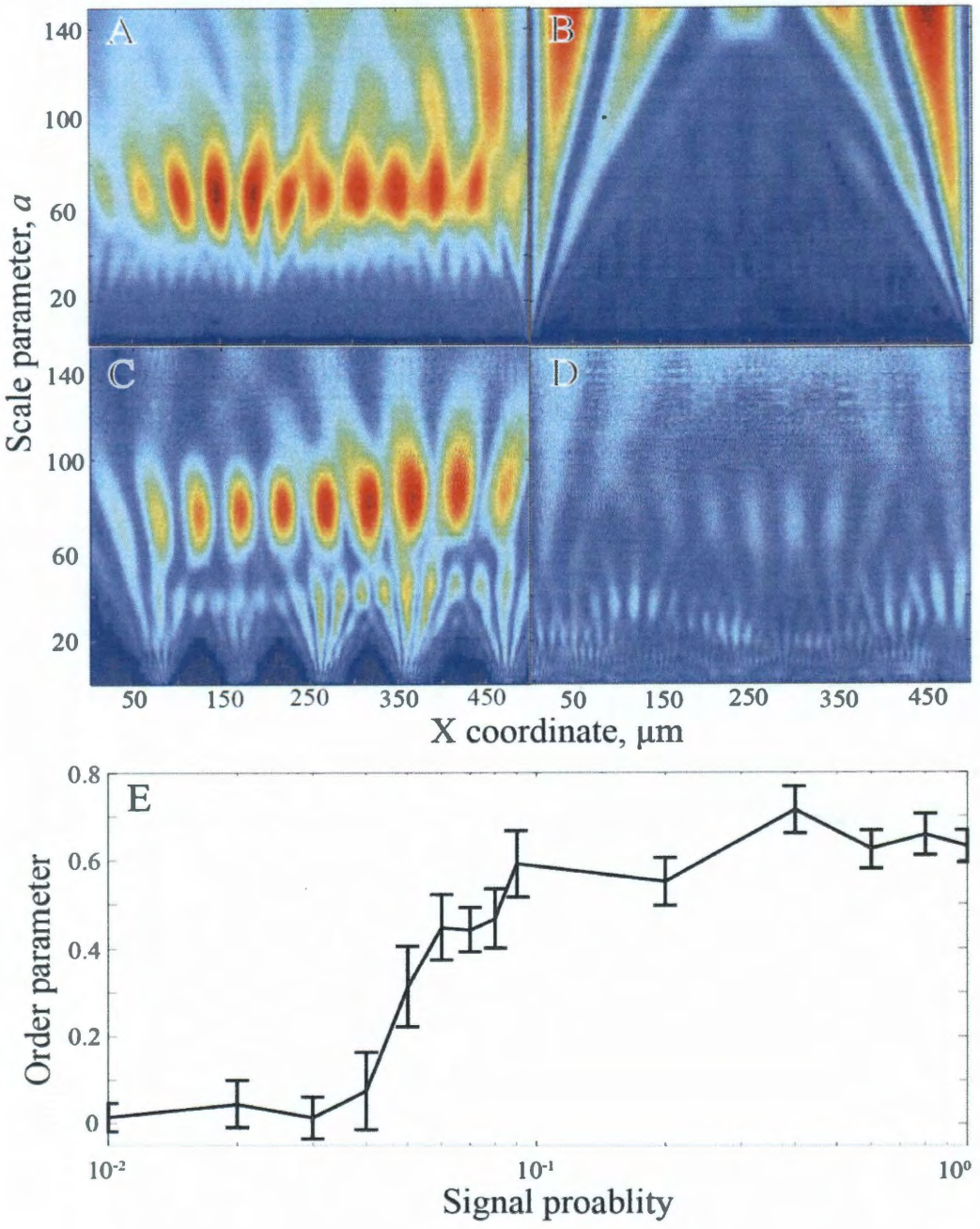
131. Jeremy, F.A. and S.W. James, *Time-Frequency Analysis of Musical Instruments.* SIAM Rev., 2002. **44**(3): p. 457-476.
132. Fontes, M. and D. Kaiser, *Myxococcus cells respond to elastic forces in their substrate.* Proc Natl Acad Sci USA, 1999. **96**(14): p. 8052-7.
133. Wolgemuth, C.W., *Force and flexibility of flailing myxobacteria.* Biophys J, 2005. **89**(2): p. 945-50.
134. Igoshin, O.A., et al., *Waves and aggregation patterns in myxobacteria.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(12): p. 4256-4261.
135. Julien, B., A.D. Kaiser, and A. Garza, *Spatial control of cell differentiation in Myxococcus xanthus.* Proc Natl Acad Sci U S A, 2000. **97**(16): p. 9098-103.
136. Kearns, D.B. and L.J. Shimkets, *Chemotaxis in a gliding bacterium.* Proc Natl Acad Sci U S A, 1998. **95**(20): p. 11957-62.
137. Kearns, D.B., B.D. Campbell, and L.J. Shimkets, *Myxococcus xanthus fibril appendages are essential for excitation by a phospholipid attractant.* Proc Natl Acad Sci U S A, 2000. **97**(21): p. 11505-10.
138. Kearns, D.B., et al., *Identification of a developmental chemoattractant in Myxococcus xanthus through metabolic engineering.* Proc Natl Acad Sci U S A, 2001. **98**(24): p. 13990-4.
139. Kearns, D.B., et al., *An extracellular matrix-associated zinc metalloprotease is required for dilauroyl phosphatidylethanolamine chemotactic excitation in Myxococcus xanthus.* J Bacteriol, 2002. **184**(6): p. 1678-84.
140. Hubbard, A.T., *Encyclopedia of surface and colloid science.* 2002, New York: Marcel Dekker.
141. Treuner-Lange, A., M.J. Ward, and D.R. Zusman, *Pph1 from Myxococcus xanthus is a protein phosphatase involved in vegetative growth and development.* Mol Microbiol, 2001. **40**(1): p. 126-40.
142. Hoiczyk, E., et al., *Lipid body formation plays a central role in cell fate determination during developmental differentiation of Myxococcus xanthus.* Mol Microbiol, 2009. **74**(2): p. 497-517.
143. Hillesland, K.L., R.E. Lenski, and G.J. Velicer, *Ecological variables affecting predatory success in Myxococcus xanthus.* Microb Ecol, 2007. **53**(4): p. 571-8.
144. Hess, B., *Periodic patterns in biology.* Naturwissenschaften, 2000. **87**(5): p. 199-211.
145. Nicola, E.M., M. Bar, and H. Engel, *Wave instability induced by nonlocal spatial coupling in a model of the light-sensitive Belousov-Zhabotinsky reaction.* Phys Rev E Stat Nonlin Soft Matter Phys, 2006. **73**(6 Pt 2): p. 066225.
146. Siegert, F. and C.J. Weijer, *Spiral and concentric waves organize multicellular Dictyostelium mounds.* Curr Biol, 1995. **5**(8): p. 937-43.
147. DeAngelis, D.L. and W.M. Mooij, *Individual-based modeling of ecological and evolutionary processes.* Annual Review of Ecology Evolution and Systematics, 2005. **36**: p. 147-168.
148. Kreft, J., G. Booth, and J. Wimpenny. *Applications of individual-based modelling in microbial ecology.* 1999.

149. Bulyha, I., et al., *Regulation of the type IV pili molecular machine by dynamic localization of two motor proteins.* Mol Microbiol, 2009. **74**(3): p. 691-706.
150. Leonardy, S., I. Bulyha, and L. Sogaard-Andersen, *Reversing cells and oscillating motility proteins.* Mol Biosyst, 2008. **4**(10): p. 1009-14.
151. Berleman, J.E. and J.R. Kirby, *Deciphering the hunting strategy of a bacterial wolfpack.* FEMS Microbiol Rev, 2009. **33**(5): p. 942-57.
152. Garcia-Ojalvo, J., M.B. Elowitz, and S.H. Strogatz, *Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing.* Proc Natl Acad Sci U S A, 2004. **101**(30): p. 10955-60.

Appendix A

Using the wavelet coefficients calculated in previous step, an order parameter is derived that describes the existence of waves. Suppose the averaged 1-D wavelet coefficients calculated using Eq. (13) of an image with no wave is $c_0(a)$ and some wavelet coefficients $c(a)$ are obtained using the same equation, the order parameter *OP* is defined as: $OP = \max(c(a) - c_0(a))$. When *OP* is smaller than 0.4, no visible ripple pattern can be seen. When OP is greater than 0.4, ripple pattern is observed.

By gradually decreasing the signal probability it is discovered that wave disappear when the signal probability reached a critical value. So a conclusion is reached that the signal probability serves as a bifurcation parameter in the agent-based model for ripples. *OP* is used as the quantification of ripple pattern and Appendix figure 1E illustrates this conclusion: the order parameter *OP* is ultrasensitive to increases in signal probability around the threshold value of 0.05 above which ripples appear.



Appendix figure 1 Signal probability is a good bifurcation parameter to control self-orgnaization into ripples. (A) The wavelet coefficient from a wavelet transform of an image contains ripple. (B) The wavelet coefficient from a wavelet transform of an image without ripple. (C) Wavelet coefficients from wavelet transform of an image with ripple from ABM simulation. (D) Wavelet coefficients from wavelet transform of an image without ripple from

ABM simulation. (E) The order parameter (see methods) is computed from wavelet coefficients as an indication of ripple presence. The order parameter is close to zero when there is no ripple and larger than 0.4 when ripple presents. The error bar is computed from 10 simulations. This plot shows that the signal probability serves as a bifurcation switch of *M. xanthus* ripple pattern.